

Every little helps

Novel techniques that could help to make human embryonic stem-cell research morally acceptable will not immediately defuse the ethical debate over the work.

This week, *Nature* is publishing online two papers describing new techniques for deriving embryonic stem cells. The limits on federal funding for human embryonic stem-cell research in the United States have inspired political lobbying and led to the creation of alternative funding sources and international networks to facilitate offshore collaborations. These papers mark a new approach: they are aimed at making human embryonic stem-cell research morally acceptable (see page 1076).

One group, from the Massachusetts-based biotechnology company Advanced Cell Technology, has used a single-cell biopsy procedure, similar to an *in vitro* fertilization (IVF) technique called preimplantation genetic diagnosis (PGD), to derive mouse stem cells without affecting the subsequent development of the donor embryo (Y. Chung *et al. Nature* doi:10.1038/nature04277; 2005). The authors propose that human embryonic stem-cell lines could be created in the same way, using embryos that are already undergoing the IVF technique as donors, so there is minimal additional risk to the embryo.

One limitation of this method is that stem cells derived in this way would represent a limited gene pool restricted to the clientele of IVF clinics, who tend to be affluent and white. But a second method, reported this week by researchers at the Whitehead Institute, also in Massachusetts, would produce cells matched to any donor.

The altered nuclear transfer (ANT) method, which has been tested in mice, switches off a gene in the donor nucleus before it is transferred to the egg (A. Meissner & R. Jaenisch *Nature* doi:10.1038/nature04257; 2005). The gene is necessary for the embryo to implant in the womb, so the embryo cannot develop beyond the implantation stage. William Hurlbut, a bioethicist at Stanford University in California and a member of the President's Council on Bioethics, argues that these embryos have neither the cellular organization nor the developmental potential of a normal embryo, and so need not be accorded the same moral status.

It remains to be seen if either method can do anything to resolve the political impasse over human embryonic stem-cell research. The PGD method was attacked in Congress earlier this year. Critics said that delaying the IVF procedure to allow one of the cells in an eight-cell embryo, or blastomere, to divide into cells that can be used for both IVF and embryonic stem-cell culture exposes the embryo to more risk than normal. Geneticists and fertilization specialists must now decide whether the idea is worth testing in humans.

The ANT method, on the other hand, is predicated on the question of when life begins — a question that scientists can't answer and that non-scientists cannot agree on. Several Harvard scientists have argued that ANT would only make an already difficult procedure even more cumbersome, and delay progress on therapies (*N. Engl. J. Med.* **351**, 2791–2792; 2004). Another question is whether there is a moral difference between making an embryo that is incapable of forming the tissue required for implantation, and stripping the tissue from the embryo after it has formed.

The ethical appeal of the approach has yet to be tested. If proponents such as Hurlbut are sufficiently numerous and influential, it could help to tip the balance towards a more benign regulatory and funding environment for human embryonic stem-cell research.

The scientific and ethical issues raised by these discoveries are complex, and by publishing these papers *Nature* is not endorsing either of them as a solution. Rather, we support the full funding of human embryonic stem-cell research using all available technologies, subject to tight ethical oversight. Now that the science has been reported, the advantages and disadvantages of both methods, as well as their feasibility in humans, can be rigorously assessed. ■

"Now that the science has been reported, the advantages of both methods can be rigorously assessed."

An unhealthy practice

Prescription guidelines should not be written by people with financial conflicts of interest.

A survey reported on page 1070 of this issue suggests that pharmaceutical companies often have financial links with the people who write guidelines on how physicians prescribe drugs. More than a third of the authors of such guidelines declared a financial connection to the companies that produced the drugs.

No one is suggesting that the authors in question are deliberately adjusting guidelines to match these interests: most of them surely have their patients' interests at heart. But that doesn't mean that the

potential conflicts don't matter. In other areas of medicine — including clinical trials and educational programmes — pharmaceutical money has been shown to exert a subtle influence on the choices made by physicians and researchers (see A. Wazana *J. Am. Med. Assoc.* **283**, 373–380; 2000). So it is not unreasonable to assume that a similar effect is at work in the preparation of prescription guidelines.

Several simple steps could be taken to improve matters. The doctors' organizations that write most of the guidelines should require that authors fully disclose potential conflicts of interest and add their disclosure statements to the guidelines. This doesn't always happen now — despite widespread agreement that it's the right thing to do.

But disclosure is only a first step towards objectivity. The survey also revealed that about 10% of all guideline panels include an author who holds stock in a relevant company. These holdings should be



restricted to modest levels, above which the authors should excuse themselves from writing guidelines. The professional societies that produce the guidelines should also consider introducing a ceiling on what authors can earn from relevant companies. Some organizations, such as the Institute for Clinical Systems Improvement, a Minnesota-based collaboration between US healthcare providers that produces advice on clinical practice, already have such criteria. There is no reason why others shouldn't follow suit.

But the time may also be ripe for a more radical overhaul of the system, particularly in the United States, where the financial stakes are highest and where the government currently plays almost no role. The organizations that produce the guidelines argue that it is impractical to exclude every individual with such a financial link from their preparation. But bodies in related spheres already do this. The Consensus Development Program at the National Institutes of Health, for example, which publishes advice on areas of clinical practice where there is debate about treatment options, operates a 'jury and witnesses' model, in which experts with financial links

can only present evidence, not make judgements on it.

A similar mechanism could be set up at the US health department, for example, to steer the preparation of clinical guidelines. This would, of course, require modest public funding. Britain's National Institute for Health and Clinical Excellence, which receives government funding and devolves its decisions to external experts, provides a useful model.

But would such an organization be able to produce guidelines quickly enough? When a new drug becomes available, a team of physicians can review the evidence and make recommendations in a few months. More sophisticated, consensus-based approaches typically take longer. But if it is adequately funded, such a body could produce guidelines expeditiously. If US physicians really want to separate marketing from medicine, they should support its creation. ■

"A 'jury and witnesses' mechanism could be set up at the US health department to steer the preparation of guidelines."

Grand ambition

Germany's coalition government is well placed to reform the country's scientific system.

In most countries, the award of a Nobel prize is a cause of much self-congratulation. But not in Germany. When Theodor Hänsch won a share of this year's physics Nobel for his work on laser-based precision spectroscopy, many commentators saw the award as an exception to the rule of underperformance in German science.

Such introspection is all too characteristic of modern Germany. The truth is that German scientific prowess and inventiveness can still compete with the best in the world in many spheres. But the science system suffers from a number of structural weaknesses. The new government, almost certainly a 'grand coalition' of the two large political parties, the Social Democrats (SPD) and the Christian Democrats (CDU/CSU), should move quickly to address them. And the science ministry, which is likely to be led by Annette Schavan, a former minister for education in the state of Baden-Württemberg, should focus on areas where intervention is really needed.

It should start by revisiting Germany's restrictive rules and regulations in stem-cell research and plant biotechnology, which make it hard for researchers in Germany to be competitive. Former chancellor Gerhard Schröder had promised to open discussions on regulations that effectively prevent research using embryonic stem cells. Schavan said she wouldn't. This is one promise she should break.

Despite its long tradition in science, Germany still lacks research universities that can compete with the best in the United States and Britain. The federal education system, which distributes strong universities all over Germany, has served the country well. But to compete internationally, Germany needs to target a few institutions and support them to the hilt. The previous government launched a €1.9-billion (US\$2.3-billion) programme to strengthen the research profile of the best German universities. The new government will need to push this programme hard if it is to overcome state rivalries

and allow the strongest research universities to grow stronger.

Across the university system as a whole, reforms are needed to ensure that appointments are made on merit. At present, political considerations and personal influence are often more important than a candidate's scientific strengths. If institute and department heads, and not ministry officials, were solely responsible for making appointments, this situation would change. Schavan says she wants the federal government to retreat from its joint responsibility for universities, and this could prove to be beneficial provided that the state (*Länder*) governments grant increased autonomy to the universities afterwards, as many of them have pledged to do.

Given a longer leash, universities could work harder to attract the best scientists from abroad. Germany has failed to make the most of the scientific talent in eastern Europe, despite its geographical proximity. While US universities rolled out the red carpet to immigrant scientists from that region, most German universities have not been flexible enough to create new positions, or to offer competitive salaries and conditions. The new government must help to put that right.

It should also take steps to make research careers more attractive for the declining number of science students in Germany itself, and should create conditions that will keep women in research. It can start by nurturing merit-based appointments and providing incentives for employers to provide decent child care (see *Nature* 437, 296; 2005). Schavan, who calls herself a 'conservative feminist', and Angela Merkel, the chancellor-designate and a former physicist, need to find ways to help more women achieve in science what they have done in politics.

Finally, money counts. Scientists will expect both coalition partners to stick to their pre-election commitments to gradually increase public science and education budgets. This also has a European dimension: it is unlikely, without Germany's active support, that the European Union's next Framework research programme will develop as it should. ■

"The new government should work to create conditions that will keep women in research."

RESEARCH HIGHLIGHTS

An allele for endurance

Hum. Genet. doi:10.1007/s00439-005-0066-0 (2005)

A single gene may help to distinguish which athletes are likely to succeed at different endurance events. The *EPAS1* gene encodes a transcription factor that is activated under low-oxygen conditions.

By comparing 492 elite endurance athletes with a control group, researchers at the University of Sydney, Australia, found that one *EPAS1* allele is more common in athletes such as swimmers and middle-distance runners, who need maximum power for periods of only minutes. Another allele is predominant in those competing in Olympic triathlon and Ironman events, who need sustained performance.

According to Ronald Trent, who led the research, the alleles may differ in how efficiently they 'switch' the body over from using mainly anaerobic metabolism — effective for short, intense energy bursts — to an aerobic energy supply, needed for prolonged periods of exercise.



CANCER

Seek and destroy

Nature Med. doi:10.1038/nm1311 (2005)

Immune cells could one day be used to hunt down and destroy solid tumours such as skin cancers. A team led by Claude Perreault at the University of Montreal, Canada, have adapted a treatment that has already been successful against human leukaemia. They primed T cells from mice with a protein fragment called H7⁺, which is found on the surface of all body cells, then injected them into mice suffering from melanoma.

Remarkably, the T cells killed the melanoma cells but did not attack healthy tissue. The team suggest that this is because the abnormal blood vessels inside tumours contain high levels of a protein called Vcam1 that attracts T cells.

calculations show that shallow-water ocean waves can be focused by an array of regularly spaced pillars on the sea bed. The effect is similar to the manipulation of light waves by arrays of microscopic scatterers in photonic crystals. The pillars act like a refractive medium; if the array has a lens-shaped perimeter, it brings incident waves to a focus. That point, say the researchers, would be an ideal place to put a device that harnesses wave power.

ZOOLOGY

Ancient sex

Proc. Natl Acad. Sci. USA

doi:10.1073/pnas.0504031102 (2005)

Placozoans (pictured) are tiny invertebrates with the simplest organization and smallest genomes of any known multicellular animal. This makes them promising model organisms for studying the development of the multicellular state. They are also completely asexual — or are they?

Ana Signorovitch and colleagues from Yale University in New Haven, Connecticut, show, in a selection of genes from ten placozoan isolates, that

the overall level of nucleotide variation is similar to the level of variation within any one individual. This is consistent with expectations for a sexual population, but not an obligately asexual one.

PATHOLOGY

Passing on prions

Science **310**, 324–326 (2005)

Kidney inflammation in prion-infected mice can cause the infectious protein particles to be excreted in their urine — suggesting a mechanism for prion transmission.

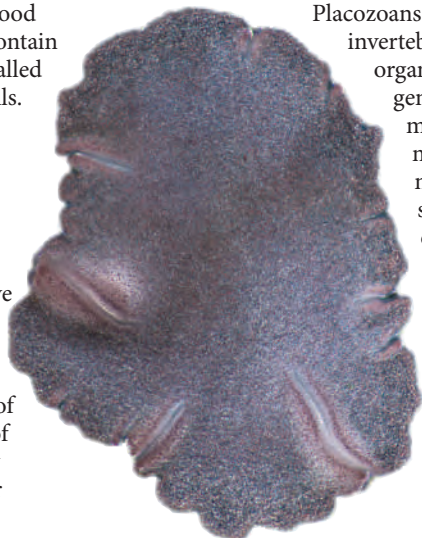
The study, by Adriano Aguzzi of the University Hospital of Zurich in Switzerland and colleagues, may explain how prions spread between sheep, causing a disease called scrapie, and give rise to a chronic wasting disease in elk and deer. It also raises concern that urine from humans with variant Creutzfeldt–Jakob disease may contain low concentrations of prions.

PHYSICS

Wave power

Phys. Rev. Lett. **95**, 154501 (2005)

“How do you keep a wave upon the sand?” a nun famously sang in *The Sound of Music*. Xinhua Hu and Che Ting Chan of Hong Kong University of Science and Technology suggest an answer. Their



CHEMISTRY

Chemical cascade

J. Am. Chem. Soc. doi:10.1021/ja055545d (2005)

Organic synthesis is a slow business if each chemical alteration to a molecule must be followed by a lengthy purification process, so chemists try to design reactions that can happen in a single cascade. A neat example comes from David MacMillan and his colleagues at the California Institute of

Technology in Pasadena.

They show that an imidazolidinone molecule can catalyse two distinct addition reactions — one electrophilic, the other nucleophilic — to build up complex products quickly from simple precursors.

This cascade reaction produces high yields and is also highly stereoselective, so that incoming chemical units preferentially add to just one face of the substrate molecule, rather than producing a random mixture of mirror-image molecules.

CELL BIOLOGY

Breaking tensions

J. Cell Biol. **171**, 153–164 (2005)

When epithelial cells become cancerous, they acquire the ability to break down the cell–cell junctions that stick them together, leaving them free to invade other tissues. These junctions are normally maintained by the protein E-cadherin.

Johan de Rooij, now at the Netherlands Cancer Institute in Amsterdam, and his colleagues have investigated a model of this process, in which scattering of cultured canine kidney cells is induced by hepatocyte growth factor (HGF).

They find that HGF does not disrupt E-cadherin function, as had been thought. Instead, the cell–cell junctions are ripped apart by tensions that build up in the cells' internal matrix, the cytoskeleton. The strength of these forces is linked to how tightly the cells adhere to the extracellular matrix, via integrin receptors.

ASTRONOMY

Watch this asteroid

Icarus **178**, 281–283 (2005)

Fortunately for us, the asteroid known as 2004 MN4 will pass close enough to Earth to allow some spectacular observations, without hitting the planet. In work that tells astronomers what to watch for, Daniel Scheeres of the University of Michigan in Ann Arbor and his colleagues predict that the asteroid's close pass will dramatically alter the way it tumbles.

Uncertainty about this asteroid's orbit created an Earth-impact scare that made headlines in December last year. But its closest approach, in April 2029, is now expected to leave roughly 30,000 km of space between it and us. This should be near enough for the rotational changes induced by Earth's gravity, simulated by Scheeres *et al.* to be visible using ground-based telescopes.



M. WILKINSON, UNIV. CAMBRIDGE

BIOMECHANICS

Winging it

Proc. R. Soc. Lond. B doi:10.1098/rspb.2005.3278 (2005)

Pterosaurs (pictured) possessed a unique extra wrist bone called the pteroid. This either pointed forwards and supported a membranous forewing, or, as an opposing theory suggests, was tucked away, alongside the arm. Researchers provide data to inform this long-running debate by testing the two wing shapes in a wind tunnel.

Matthew Wilkinson and colleagues from Cambridge University, UK, made scale models of pterosaur wings with and without a forewing supported by a pteroid bone; more lift was obtained with the forewing. They claim that this settles the argument, and explains how even the largest pterosaurs could take to the air.

SPINTRONICS

Caught on a jewel

Nature Phys. doi:10.1038/nphys141 (2005)

Impurities in diamond may dismay the jeweller, but they excite the information engineer. A defect called a nitrogen-vacancy centre, created by impurities in the diamond, has an electron spin that could be used to encode data, possibly for quantum computing.

David Awschalom and his colleagues at the University of California, Santa Barbara, have developed an instrument that combines magnetic and optical probing to measure the spin of a single nitrogen-vacancy defect. By monitoring this site, they are also able to detect interactions with neighbouring impurities, including 'dark' spins, which are invisible to conventional detection by photoluminescence.

JOURNAL CLUB

Bob Behringer

Duke University, North Carolina, USA

A grain of a good idea, introduced by a material scientist.

Imagine a handful of sand. With a bit of force, you can push it around. But what is the nature of the deformations occurring within the sand, and how do they lead to its macroscopic behaviour?

For some time now, I have been fascinated by the behaviour of granular materials. These materials share a common feature with disordered molecular solids, foams and other many-body systems: they exist in jammed states, which deform irreversibly when a finite force is applied.

Craig Maloney and Anaël Lemaître, then both at the University of California, Santa Barbara, published two papers in *Physical Review Letters* addressing how the deformation happens (93, 195501 and 016001; 2004).

Their work uses numerical and physical modelling of the microscopic behaviour of particles from which granular-type materials are made. They consider deformations of granular matter using shear transformation zone theory, an approach that had previously been applied only to molecular solids.

Several testable predictions emerge from their observations, including that plastic (irreversible) changes in the material structure should alternate with elastic (reversible) deformation, and that the elastic constants will show unusual behaviour. These predictions can be put to careful experimental test.

And here is where the idea becomes really appealing. Granular experiments of the type performed by my group (T. S. Majumdar & R. P. Behringer *Nature* **435**, 1079–1082; 2005), unlike molecular studies, can directly access the particle dynamics. If this approach proves valid, granular materials may well inform us about processes at the molecular scale.

NEWS

Cash interests taint drug advice

Researchers and physicians who write the rules on prescribing drugs have extensive financial connections with the pharmaceutical industry, an investigation by *Nature* has revealed. Public-health experts say that the results of the survey, which is the largest of its kind, suggest that drug companies are distorting decisions about how their products are being prescribed.

In the investigation of the panels that write clinical guidelines — documents that govern the diagnosis and treatment of patients — *Nature* found that more than one-third of authors declared financial links to relevant drug companies, with around 70% of panels being affected. In one case, every member of the panel had been paid by the company responsible for the drug that was ultimately recommended.

These links with pharmaceutical companies are more worrying than the financial conflicts known to plague clinical trials and reviews, say public-health experts, because the guidelines have such a direct effect on the drugs that doctors prescribe. “The guidelines are specifically written to influence the practice of many physicians,” says Niteesh Choudhry, a specialist in health policy at Harvard Medical School. “The effects of conflicts may be translated many times over to patients.”

“The numbers in the survey are distressing,” agrees Drummond Rennie, deputy editor of the *Journal of the American Medical Association* and an advocate of proposals to free guidelines from industry influence. “Drug-company sponsors see guideline-issuing bodies as perfect places to exert influence. The practice stinks.”

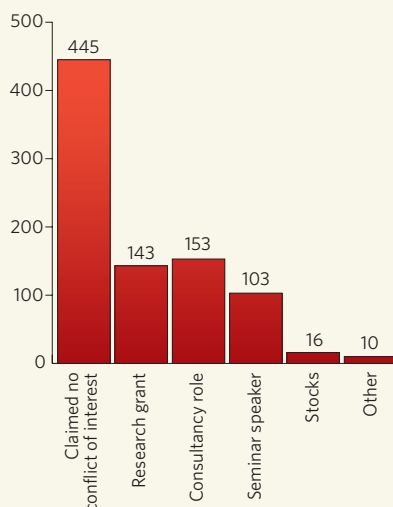
Misguided?

The guidelines have moved on to doctors' radar screens over the past decade or so, in line with an increased emphasis on using ‘evidence-based’ methods to treat patients. Many major professional bodies around the world, such as the American College of Physicians, now publish several guideline documents every year. The advice is based on the results of clinical trials, and tells doctors which drugs to prescribe and when.

In its survey, *Nature* studied more than 200 guidelines from around the world that were deposited with the US National Guideline Clearinghouse in 2004. Only 90 contained details about individual authors' conflicts of interest. Of those, just 31 were free of industry influence.

CONFLICTS OF INTEREST

In 685 disclosures examined in *Nature*'s survey of authors of prescription guidelines.*



35% of authors said they had a conflict of interest of some kind.

16 authors helped to write guidelines on illnesses relevant to companies in which they owned stock.

49% of guidelines did not include any details of authors' conflicts of interest.

For full survey results, see www.nature.com/news/2005/051017/full/4371070a.html

* Some of the authors had more than one conflict of interest

Half of those that gave individual details had at least one author with a conflicting advisory position in the pharmaceutical industry. More than a third of the panels included at least one member who gave seminars on behalf of a relevant drug company. And one in ten contained a member who owned stock in a company with products being considered.

The latter type of conflict is especially worrying, says Bruce Fye of the Mayo Clinic in Rochester, Minnesota, who raised the issue of guideline ethics during his presidency of the American College of Cardiology in 2002–3. Prescription practices promoted by guidelines can have a direct effect on drug sales and thus stock price. “Stock options should be viewed as unacceptable,” he says.

“It's not true that people can remain completely objective.”

In one example uncovered by *Nature*, guidelines for the treatment of anaemia in HIV-positive patients were written by a working group selected by Paul Volberding, a leading AIDS researcher and physician, and vice-chairman of the Department of Medicine at the University of California, San Francisco. Volberding convened the group at the request of Ortho Biotech, a pharmaceutical company based in Bridgewater, New Jersey. Ortho Biotech funded the group's meetings, and all six members, including Volberding, had been paid by the company for lecturing or consultancy jobs. The group's latest guidelines, published last year (P. Volberding *et al. Clin. Infect. Dis.* **38**, 1454–1463; 2004), recommend the use of epoetin alpha, a drug marketed by Ortho Biotech.

Slippery stats

Physicians' organizations say that one or two authors with a conflict of interest could not influence a panel containing tens of members. But assessing just how many authors have such conflicts is difficult. In the *Nature* survey, 35% of the 685 authors involved declared conflicting industry links (see graphic). Experts say that this is likely to be an underestimate, however, because it relies on authors' own declarations.

For example, the Center for Science in the Public Interest (CSPI), based in Washington DC, examined the disclosure statements on some randomly chosen blood-pressure guidelines published last year. Only one author declared a conflict of interest, but further checks by the CSPI revealed that four other authors had failed to disclose research funding from relevant drug companies.

The problem could be even worse in guidelines that don't contain conflict-of-interest declarations, warns Merrill Goozner, director of the CSPI's Integrity in Science project. “It is usually the journals and supplements that rely heavily on industry advertising that are least likely to have good disclosure policies.”

The organizations that produce guidelines argue that industry links do not necessarily lead to biased advice. Some employ review panels to check prospective guidelines — although conflicts of interest are just as likely to occur in review panels. And authors insist that they are careful to avoid bias. “We realized this would be a concern and so looked very

**CONFERENCE COVERAGE**

News from the Geological Society of America and the American Society for Reproductive Medicine.

www.nature.com/news

carefully at our recommendation,” says Volberding of the HIV guidelines. “The committee really was independent in its deliberations.”

Critics say that the influence exerted by industry money is unconscious but powerful. Funding for conference travel and accommodation has been linked with increased requests for a sponsor’s drug, for example. And meetings with drug reps both decrease physicians’ ability to spot erroneous claims about medications and increase their requests for new drugs with no demonstrated advantage over generics (A. Wazana *J. Am. Med. Assoc.* **283**, 373–380; 2000).

“It’s not true that people can remain completely objective,” says Peter Gøtzsche, director of the Nordic Cochrane Centre in Copenhagen, Denmark, which runs independent reviews of drug efficacy.

Gøtzsche and others want guidelines to be prepared in ways that reduce the potential for bias. Some such schemes already exist. In Britain, for example, guidelines are provided by a government-funded body called the National Institute for Health and Clinical Excellence (NICE). Although some of the experts used by NICE have industry links, the institute itself is independent of industry and consults with other stakeholders, such as patients’ groups.

Expert witnesses

In the United States an alternative process is being used for conditions where there is conflicting evidence as to the best treatment. Experts in a particular condition — those most likely to have relevant industry links — are used as witnesses, but the panel members themselves are chosen for their skills at examining medical data. “If you have a conflict of interest you can’t be on the panel,” says Susan Rossi, deputy director of the Office of Medical Applications of Research at the National Institutes of Health, which runs the scheme.

But the bodies that produce guidelines maintain that there just aren’t enough experts without conflicts of interest. Nathaniel Clark of the American Diabetes Association estimates that three-quarters of members eligible to write guidelines have industry links, and other organizations report a similar number.

The latest figures underline the urgency of breaking the deadlock. In 2002, Choudhry and colleagues found similar levels of conflicts of interest, but he says that physicians have since assumed that disclosing conflicts is all that is required to solve the problem. With drug-company stock-holders still sitting on the panels that make prescription decisions, Choudhry warns: “We need a better way to deal with conflicts of interest.” ■

Rosie Taylor and Jim Giles



Europe’s first mission to Venus is set for launch later this month.

ESA

Express delivery to Venus

Get set for a revival of interest in Earth’s wayward twin. The European Space Agency (ESA) plans to launch Venus Express next week — the first visit to the planet in more than a decade. The focus of the €220-million (US\$265-million) mission will be the planet’s thick atmosphere, which some think could host exotic life forms that use ultraviolet instead of visible light.

Venus Express will begin its journey from Kazakhstan on a Soyuz–Fregat rocket; the month-long launch window opens on 26 October. ESA developed the mission — its first to Venus — in less than four years, after asking the science community for ideas on adapting its Mars Express spacecraft for another purpose.

The craft uses cameras and spectrometers adapted from Mars Express and Rosetta, an upcoming comet mission. These will build up a three-dimensional profile of Venus’s carbon dioxide-rich atmosphere, which traps heat and drives temperatures on the surface to a searing 450 °C. Past missions, including the 1978 Pioneer Venus orbiter and the Magellan surface-mapper of the 1990s, have hinted at active volcanism and lightning on the planet; Venus Express will investigate these observations further.

The mission may also help to determine what is absorbing ultraviolet radiation in the planet’s clouds. David Grinspoon of the Southwest Research Institute in Boulder, Colorado, and others speculate that life may have evolved that thrives in acidic environments and photosynthesizes

ultraviolet light. Rather than living on the hot surface, such organisms might reside in long-lived clouds, where temperatures and pressures are more Earth-like.

Grinspoon has been modelling the history of Venus’s atmosphere, and his preliminary results indicate that the planet’s early ocean could have persisted for as long as 2 billion years. Life could have adapted to the clouds when the surface water disappeared, he suggests.

Grinspoon calls this a kind of “thought experiment”, and does not claim there is evidence for life on Venus today. Nor does he believe that terrestrial microbes stowing away on Venus Express risk contaminating the planet. Current international guidelines for planetary protection require sterilization for Mars-bound spacecraft, but not for those headed for hot and acidic Venus.

Still, NASA has asked the US National Academy of Sciences, based in Washington DC, to revisit the subject of planetary protection for Venus. Thirty years have passed since it last looked at the question, says NASA planetary-protection officer John Rummel. At present, the agency is mulling over ideas for sample-return missions to Venus, and the academy has recommended that a lander be sent to the planet in the next decade.

In the meantime, Japan plans to launch its Venus Climate Orbiter in 2008, and two Mercury-bound spacecraft — MESSENGER and BepiColombo — will study Venus during brief fly-bys. ■
Tony Reichhardt

Quake aid hampered by ban on web shots

Open-access satellite images are revolutionizing responses to disasters. Yet the government of Pakistan has forced aid agencies to remove pictures of earthquake devastation from the Internet.

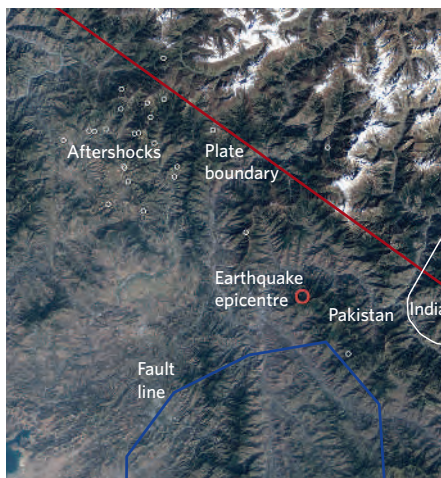
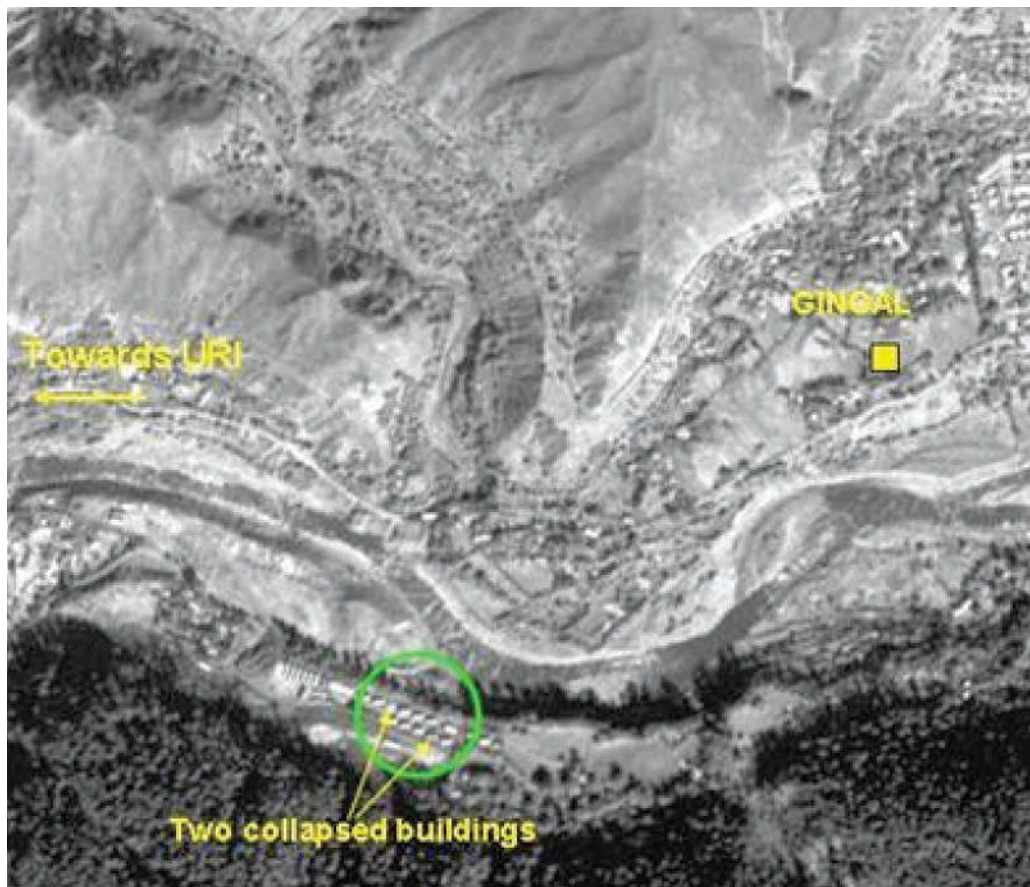
Three days after the 7.6-magnitude earthquake struck Kashmir on 8 October, the Pakistan government appealed for high-resolution satellite images to help relief efforts. But, apparently to protect national security, *Nature* has learned that the government has since forced international agencies and relief organizations to remove these images from their websites.

The International Charter on Space and Major Disasters put high-resolution images of the earthquake zone on its website last Friday, then pulled them off hours later. The charter, a consortium of space agencies, was created in 2000 to supply satellite images and data to communities in need of relief following a disaster.

An International Charter spokesperson said: "To best aid relief efforts, we are no longer publicly disseminating pictures of the Pakistan earthquake. Publication of such images would compromise the ability of United Nations (UN) forces on the ground to deliver relief. We hope you understand the situation."

But a senior official at the charter, who asked not to be named, says that the Pakistan government had demanded that no photos be made accessible to the public, because it feared the images could compromise security in the Kashmir region — an area that has long been disputed territory between India and Pakistan. The UN and other aid agencies need Pakistan's cooperation on the ground, and had no choice but to comply, he says.

The UN, European Union (EU) and other international agencies have a general policy of making all such images publicly available. But



Images of the devastation in Kashmir caused by the earthquake in Pakistan could help direct relief efforts, but high-resolution shots of the area (above) were removed from public websites.

be named. The images are useful to other local aid organizations, and to professionals worldwide who help out using the Internet, the official points out.

Scene from above

The Citizens Foundation, for example, a well-respected Pakistan organization based in Karachi, is providing basic care packages including tents, blankets and food rations to those affected. But Ayaz Abdulla, a mechanical engineer from Karachi who coordinates Internet-based parts of the local relief effort, and others at the organization, spent much of last week desperately e-mailing space agencies and commercial suppliers, to try to obtain images.

"We look for major areas of dense devastation that correspond to higher population, and for access roads to allow relief-goods transport," Abdulla says. The images are used to plan logistics, work out what roads are open, and locate isolated settlements.

The importance of publicly available satellite images and other geographical data was

last week photos of the quake zone disappeared from sites such as the UN's ReliefWeb and satellite imaging site UNOSAT, and Reuters' AlertNet. On others they became password-protected, such as on the EU's disasters site. Images of the earthquake zone will now be distributed on a "need to know" basis to affiliated relief organizations, says a senior EU official.

Although the main organizations can still access the data, the pulling of the images is "regrettable", says the official, who asked not to



highlighted after the Indian Ocean tsunami in December 2004 and this year's hurricane Katrina. In the aftermath of the tsunami, thousands of geographers and researchers provided high-quality information on the Internet, "while government agencies were still struggling to call people back from their holidays", says Thierry Rousselin. A geographical-information-systems consultant based in Paris, Rousselin monitored how satellite images were used after the Indian Ocean tsunami.

The response to Katrina was even greater, partly because of the good quality and availability of public geographic data for the United States, and because the government was perceived to have made a poor job of relief efforts. Concerned web users quickly downloaded images of New Orleans, and published them using browsers such as Google Earth and Google Maps, both of which allow users to zoom in on any part of the globe to view satellite images and associated local information. Anyone with a web browser could then add information about the needs in a particular location to these webmaps, and thousands did.

Four days after the hurricane hit, a collaboration called the Global Connection realized how useful such information would be to relief

workers. The Global Connection consisted of Google and scientists at Carnegie Mellon University in Pittsburgh, Philadelphia, and NASA Ames Research Center in Moffett Field, California. Within hours they had added thousands of near real-time aerial and satellite images of New Orleans into Google Earth and Google Maps. Internet users then converted these images into a myriad of web services allowing anyone to report, for example, damage, deaths or needs at particular locations.

Out of range

High-resolution satellite images, combined with tools such as Google Earth and Maps, are "bringing a revolution", Rousselin says. "They have advanced remote sensing more in 6 months than professionals have done over the past 30 years."

Even with the limited data available for Pakistan, Internet users have acted faster in mapping the area than government and international agencies, Rousselin argues. On the morning after the Pakistan earthquake had hit, Internet users worldwide were already busy plotting the disaster in Google Earth ahead of official agencies, he says.

But Anne Wright, a computer scientist at Ames Research Center who was involved in mapping the damage after Katrina, says the lack of images of Pakistan is hampering the Global Connection's efforts. "We haven't played a similar role with the Pakistan earthquake because we've had no data sources to work with. After Hurricane Katrina, the National Oceanic and Atmospheric Administration made their post-hurricane imagery publicly available. For Pakistan, the only available imagery we're aware of is commercial satellite photography."

She and others have been trying to help the relief groups that are seeking images. Google has now purchased some imagery from DigitalGlobe, a database of high-resolution earth-imaging satellite pictures, and has provided this as photographic overlays that anyone can import into Google Earth.

But the lack of data is still being sorely felt on the ground. One of Abdulla's colleagues at the Citizens Foundation, who did not wish to be named, pleaded in an e-mail to NASA and others: "Could you please tell me whether it's possible to get updated satellite imagery for this region (northern Pakistan), just like it was after Hurricane Katrina? Our country needs all the help it can get, and much more." ■

Declan Butler

As *Nature* went to press, the UN informed relief agencies that it would lift its ban on the public dissemination of images. For further updates see:

♦ www.nature.com/news

NRSA, COURTESY OF INT. CHART SPACE MAJOR DISAST.



**SEISMOLOGIST DELAYED
EN ROUTE TO PAKISTAN**
Quake researcher fights for access to rare data in the stricken Himalayas.
www.nature.com/news

B. K. BANGASH/AP

Biologists forced to reassess embryo test

MONTREAL

A technique for detecting genetic abnormalities in embryos is itself coming under the microscope. Preimplantation genetic diagnosis (PGD) is used in fertility treatment to check the condition of an embryo before it is implanted in a mother's womb. But studies now suggest that there is still much to learn about the procedure.

The findings suggest that genetically 'abnormal' embryos could still give rise to normal embryonic stem-cell lines. And they lend urgency to a newly launched effort to track the safety of the procedure, which is becoming increasingly popular but is regulated differently around the world.

In PGD, a cell is taken from the embryo at the eight-cell stage and tested for genetic abnormalities. If the cell is given the all-clear, the embryo is implanted in the mother's uterus. Babies born after the procedure seem to be fine, but few data have been collected on the children later in life.

It is also difficult to know how well the technique works in different cases. PGD is currently targeted at older women trying to have a baby. But whether it actually cuts the rate of miscarriage among these prospective mothers is only now being examined, researchers told a joint meeting of the American Society for Reproductive Medicine and the Canadian Fertility and Andrology Society in Montreal this week.

And results presented at the meeting show that PGD picks up surprisingly high rates of

genetic abnormalities in younger women who donate eggs to infertile couples. That led some specialists, such as Jeffrey Nelson of the Huntington Reproductive Center in Pasadena, California, to suggest that the technique should be used more widely. "You can make an argument that PGD should be done in younger women", not just older women, he says.

But the most intriguing data presented at the conference suggest that embryos diagnosed as abnormal might be able to correct their genetic defects as they grow. A team including Santiago Munné of Reprogenetics, a company specializing in PGD based in West Orange, New Jersey, grew 55 embryos previously diagnosed as abnormal to the preimplantation stage, called the blastocyst. They found that many of the embryos' cells were genetically normal. One embryo contained 76% normal cells after 12 days in culture. On average, 48% of the embryos' cells were genetically normal by the blastocyst stage.

These results hint that such embryos might be an 'ethical' source of embryonic stem cells for use in research. Normally, producing these cells means destroying a viable embryo, leading many groups to brand the practice as unacceptable (see page 1076). But the genetic abnormalities of embryos rejected by PGD mean that they are unlikely to develop into a healthy baby. "These results suggest that we

could get normal stem cells from abnormal embryos," says Munné. "These embryos will never implant, so this is ethically the best way of making them."

But the results also raise questions about whether cells that are not directly sampled by PGD are actually defective. If they aren't, the potential of these embryos remains unclear. "These are the really important data we have to get from PGD," says Paulette Browne, a reproductive endocrinologist at the Shady Grove Fertility Center in Virginia. "It could be that

some of the cells are abnormal and some are not, and we're starting to collect the rest of those data right now."

Together, the latest studies highlight the importance of a new effort to accumulate data

on PGD, described at this week's meeting. Scientists and policy analysts announced the creation of a US database to track the safety of the procedure. A similar database exists in Europe, where PGD is tightly regulated, but US clinics, which carry out thousands of procedures, are not required to report their data.

The US database, set up by the Johns Hopkins University's Genetics and Public Policy Center in Washington DC, will collect information on who is using the technique in the country, the reasons for its use, and whether it results in healthy children.

Erika Check

"These results suggest that we could get normal stem cells from abnormal embryos."

China launches plans for space exploration as taikonauts touch down

The crew of China's second manned space mission returned safely to Earth on 16 October, boosting the country's reputation as a spacefaring nation.

"This proves they are now a player to stay," says Joan Johnson-Freese, an expert on China's space programme at the US Naval War College in Newport, Rhode Island. Economic development, education and national pride are important motivating factors, but the mission also serves as a global advertisement of China's technological capability, Johnson-Freese says.

The launch follows China's first manned mission in October 2003, when Yang Liwei spent 21 hours in

**IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS**

Fei Junlong (left) and Nie Haisheng arrive back from their five-day trip in space.

orbit. This mission was more ambitious, with two 'taikonauts' — Fei Junlong and Nie Haisheng —

spending five days in orbit in their Shenzhou craft, which is a close copy of the Russian Soyuz capsule. During the mission, the taikonauts took off their space suits to conduct undisclosed scientific experiments in a habitation module attached to their small re-entry capsule.

Shortly after they landed in Inner Mongolia, the director of the Manned Space

Engineering Office, Tang Xianming, outlined plans for the next mission. "Our estimate is that around 2007

our astronauts will walk in space," Tang told the state-run newspaper *China Daily*.

Chinese delegates to an international conference on lunar exploration in Toronto, Canada, last month also outlined near-term plans for robotic missions to the moon, including the Chang'E scientific orbiter set to launch in 2007. "China will be active in humanity's exploration of unknown space," Hao Xifan of the Beijing-based China National Space Administration told the conference. He did not comment on Chinese plans to send astronauts to the Moon.

Mark Peplow and
Tony Reichhardt

AFP/GETTY IMAGES

ON THE RECORD

“The American prawn cocktail is great. I've had worse in a number of restaurants on the ground.”

Gregory Olsen, a New Jersey businessman who paid US\$20 million to get aboard the International Space Station, describes the reconstituted seafood he ate while in orbit.

“We are the world's experts on hurricanes, but we're desperate. We need help.”

Mike Black of the National Oceanic and Atmospheric Administration bemoans the lack of investment and out-of-date equipment in the US hurricane forecasting department.

Source: AFP, Miami Herald

SCORECARD**Mars**

A webcam that beams images 'live' from Mars

has been activated, allowing armchair enthusiasts to sneak a peek at the red planet.

► <http://themis.asu.edu>

**Shrubs**

Toyota in Japan has come up with an attractive

method to clean the air. The car maker has developed a genetically modified version of cherry sage, which absorbs pollutants and sprouts pretty pink blossoms.

**Language**

Indigenous people in the Arctic have come to

terms with climate change, deciding that they need to add a word to the Inuktitut language to describe the phenomenon.

NUMBER CRUNCH

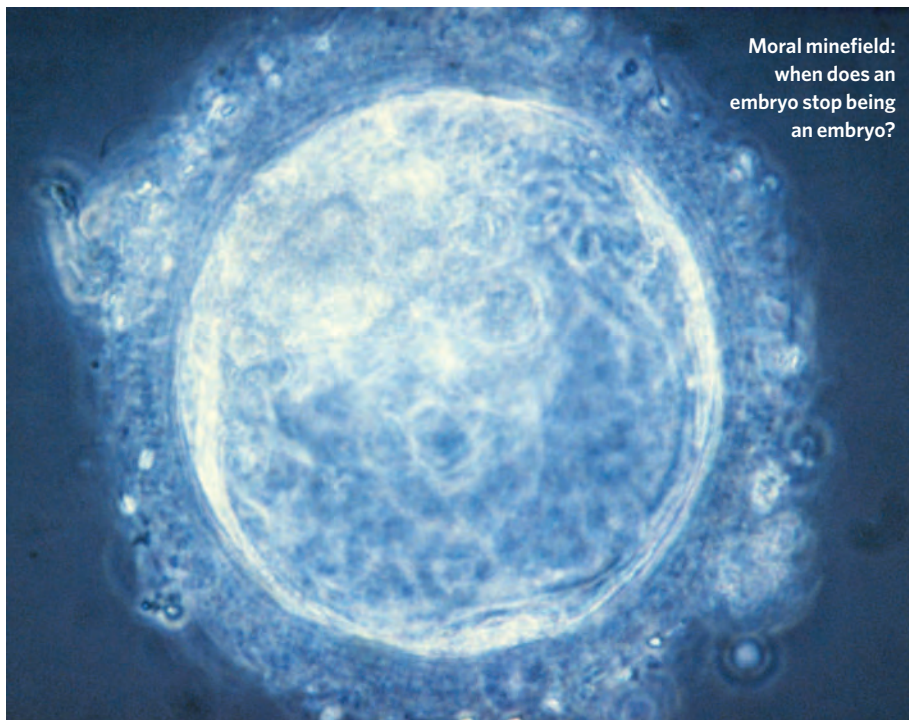
A significant part of the human genome seems to be patented...

23,688 human genes are in the database of the US National Center for Biotechnology Information.

4,382 of those genes are covered by patents.

63% of those patents are owned by private companies.

Source: K. Jensen and F. Murray Science **310**, 239–240 (2005).



Moral minefield:
when does an
embryo stop being
an embryo?

A. WALKER, MIDLAND FERTILITY SERVICES/SPL

'Ethical' routes to stem cells highlight political divide

Religious and ethical concerns are forcing researchers using human embryonic stem cells to seek ways to sidestep these issues. In a first attempt, biologists this week revealed details of two techniques for deriving the cells that do not involve the destruction of a viable embryo.

Both methods work in mice and, in principle, could be applied to human embryos. But scientists, ethicists and politicians are split over the merits of the two techniques.

Embryonic stem (ES) cells can develop into any tissue in the body, so have great potential for a range of therapies. And if they are cloned from a cell in the patient's body, the resulting cell line would genetically match that individual. This can be done using a technique dubbed 'therapeutic cloning', in which the nucleus from the donor cell is inserted into an egg stripped of its own genetic material. Earlier this year, researchers in South Korea became the first to use this method to derive human ES cells that genetically matched the patient (see *Nature* **435**, 393; 2005).

But the ethical concerns inherent in destroying an embryo to create the cell lines are too much for some, especially religious groups, and

many countries have restricted the research that can be done. In Germany, for example, the use of ES cell lines created since January 2002 is illegal and carries a penalty of up to three years in jail. And in the United States, federal funds for ES cell work is only available for a handful of cell lines derived before August 2001.

But there is currently heated debate on the issue in the US Senate, where lawmakers are considering loosening the restrictions. Some senators unhappy with those proposals have suggested that 'alternative' methods of deriv-

"You have the embryonic equivalent of brain death."

ing the cells, which don't require the destruction of viable embryos, could help to bridge the ethical divide (see *Nature* **436**, 309; 2005).

Until now, such methods have been purely theoretical, but in work published online by *Nature* this week, two teams report their successful use in mice. Rudolf Jaenisch and Alexander Meissner of the Massachusetts Institute of Technology describe a variant of therapeutic cloning called altered nuclear transfer (ANT), in which a gene in the patient's donated cell is switched off before the nucleus is transferred into a fertilized egg. The resulting egg grows into a normal ball of cells called a blasto-



STEM CELLS IN FOCUS
Catch up on all of Nature's stem-cell coverage at:
www.nature.com/news/infocus/stemcells.html

cyst from which ES cells can be derived, but the deactivated gene means that the ball lacks the ability to implant in a uterus and so develop into a baby (A. Meissner and R. Jaenisch *Nature* doi:10.1038/nature04257; 2005).

In the other paper, a team led by Robert Lanza of Advanced Cell Technology in Worcester, Massachusetts, plucked single cells called blastomeres from eight-cell embryos. They derived new ES cell lines from the blastomere, while the embryos went on to form apparently healthy mice (Y. Chung *et al. Nature* doi:10.1038/nature04277; 2005).

This method is similar to a technique used in *in vitro* fertilization (IVF) called preimplantation genetic diagnosis (PGD), in which a blastomere is removed from the eight-cell embryo for genetic tests before it is implanted. The work by Lanza's team raises the possibility that fresh stem-cell lines could be derived from human embryos being used in IVF before they are transferred to the uterus.

Although the quality of the work is impressive, there is much disagreement over the ethical benefits of each strategy. Some scientists seem more convinced by the PGD method.

In the balance

"In my mind, this takes away the ethical dilemma of destroying embryos," says Alan Trounson, a reproductive biologist at Monash University in Melbourne, Australia. There is currently a moratorium on therapeutic cloning in Australia, due for review at the end of this year, and Trounson has applied to the country's major medical funding agency to do similar work with human PGD embryos.

But there is the practical disadvantage that the resulting cell lines can come only from the embryos of couples undergoing IVF, so wouldn't be genetically matched to patients. And ethicists are troubled by the question of whether the extracted blastomere itself has the potential for life. "If you grow it in certain conditions, it could divide and differentiate to have the same properties as embryos," explains Yuri Verlinsky, chief executive of the Reproductive Genetics Institute in Chicago.

Although the number of successful births from PGD embryos indicates that removing a single cell early on doesn't compromise the baby, it is still possible that it might have subtle long-term consequences (see page 1075). "You are getting a live birth, but are you getting the same child you would otherwise get?" asks William Hurlbut, a consulting professor at Stanford University and a prominent advocate

of ANT. "It is uncomfortable to me to endorse such a strategy."

Because of this, Hurlbut says that the PGD method is unlikely to get past the Dickey Amendment, which is passed by the US Congress every year and forbids federal funds being spent on experiments that endanger or destroy an embryo. And President George W. Bush's Council on Bioethics, on which Hurlbut sits, dismissed the idea earlier this year in a white paper on alternative means for deriving ES cells.

The ANT method of altering the genetic make-up of an embryo troubles some scientists, but seems more acceptable to conserva-

tive ethicists and religious figures. "I think this is an artificial concept and I'm not comfortable with it," says Trounson. "You do an engineering step to essentially destroy the embryo so that you can then use it." Because

of this, some argue that ANT might itself fall foul of the Dickey Amendment.

George Daley of Harvard Medical School in Boston is also unconvinced, partly because the effects of the genetic modification don't kick in until the eight-cell stage. "A normal embryo and the embryo created by this method are indistinguishable until that stage," he says.

But Hurlbut maintains that the altered embryo has no moral status. "You have the embryonic equivalent of brain death," he says. "This changes the dynamic of the political debate," agrees analyst Eric Cohen of the Ethics and Public Policy Center in Washington DC.

Hurlbut is doing all he can to push ANT, and has compiled a public letter supporting it signed by scientists, ethicists and religious figures. And in August 2004, he persuaded William Levada, one of the most prominent Catholics in the United States, to write to President Bush encouraging him to consider the method. The president's bioethics council also gave the idea tentative support in its white paper.

Some observers warn against overreacting to the work. "If science gets us to the point where we don't need embryos any more that's fine, but right now policy-makers are making a huge mistake if they say 'we've got one paper and we'll make policy based on that,'" says Sean Tipton of the Coalition for the Advancement of Medical Research in Washington DC.

And either way, the ethical debate over what constitutes life — or the potential for life — looks set to dog the field. "The challenge is to define what an embryo is," says Hurlbut. "We need to sort that out or we'll be having this argument all the way along."

Carina Dennis and Erika Check

See also **News & Views** doi:10.1038/nature04305

Korea launches network to share cloning information

The World Stem Cell Hub, an international network for exchanging embryonic stem-cell lines and cloning technology, has been launched by the South Korean government.

Unveiled on 19 October, the hub will be headed by Woo Suk Hwang, who shot to international fame last year for successfully deriving human embryonic stem-cell lines by therapeutic cloning.

The hub's headquarters will be at Seoul National University Hospital, but it will have branches around the world that will train researchers in the technique, provide a bank of cell lines and, where local laws permit, create patient-specific lines. The first branches will be in Britain and California, but Hwang told *Nature* that he is also talking to researchers in Spain, Sweden and France.

Organizers hope that the first regional branch will be open by January. Each branch will need to find its own funding, but South Korea will establish a non-profit foundation to support the hub's headquarters and the travel of Korean technicians to foreign sites. Gerald Schatten, a reproductive biologist at the University of Pittsburgh, Philadelphia, who will chair the hub's international board of directors, says that the South Korean government is providing about US\$50 million.

Western scientists have cautiously welcomed the development. "I'm pleased that the Koreans have been as willing as they have to share their technology," says Arnold Kriegstein, director of the Institute for Stem Cell and Tissue Biology in San Francisco, whose staff have visited Hwang's lab.

"The Koreans are the experts — no one approaches their efficiency," adds Stephen Kennedy, a clinical researcher in reproductive medicine at the University of Oxford, UK.

But Kriegstein has ethical concerns about egg donations, and the associated issues of informed consent and record-keeping. He says he is also worried about the technology becoming centralized at such an early stage. Michael German, a diabetes researcher at Kriegstein's institute, agrees: "I would not like to see it become a specialized club where only a limited number of scientists have access to the technology."

Carina Dennis

Retractions trigger lawsuit over a damaged reputation

A legal battle has broken out over two retracted plant-biology papers.

In a lawsuit filed on 26 August, Meena Chandok of the University of Maryland, Baltimore, alleges that her former postdoctoral supervisor and co-author damaged her scientific reputation by retracting the papers without her consent. The papers appeared in 2003 and 2004 in *Cell* (M. R. Chandok *et al.* *Cell* 113, 469–482; 2003) and *Proceedings of the National Academy of Sciences* (M. R. Chandok *et al.* *Proc. Natl Acad. Sci. USA* 101, 8239–8244; 2004). They concerned nitric oxide, a chemical that is increasingly being recognized as a major signalling molecule.

Chandok's former supervisor, Daniel Klessig of Cornell University in Ithaca, New York, made the retractions late last autumn together with his other co-authors, saying they were not able to replicate some of the key data. Press reports also state that Klessig and Chandok had disagreed over how long she would work in his laboratory. A date for the hearing has not been set.



High hopes: CryoSat launched on 8 October, but fell into the Arctic Ocean shortly after.

Crashed polar satellite could fly again, says ESA

The European Space Agency (ESA) hopes to rebuild CryoSat, its crashed Earth observation mission. A new launch of the satellite, which would monitor changes in the thickness of the polar ice sheets, could take place by 2008.

CryoSat crashed into the Arctic Ocean shortly after launch atop a converted Russian missile on 8 October. It was intended to be the first project in ESA's Earth Explorer

programme, a series of satellites designed to collect data on global environmental issues.

Building and launching a 'clone' of the crashed satellite could cost less than the €136 million (US\$164 million) spent on the original mission, says Volker Liebig, director of ESA's Earth observation programmes. Whether a rebuilt satellite would use the same Russian launcher depends on the results of an inquiry into CryoSat's loss, he adds.

Radical moves proposed to help US keep its global edge

A report from the US National Academies recommends a slew of measures to ensure that the United States remains a leader in global research.

With scientific output from Asia booming, the agency wants the federal government to respond by spending an extra \$10 billion a year on education and research. Suggested schemes include 10% annual funding increases in basic research in the physical sciences, engineering, maths and information sciences for the next seven years; a new agency for energy research; and a one-year visa extension for international researchers,

S. CORVAIA/ESA

who would receive automatic work permits if they found a job.

The report, which was requested by Congress, lists many examples where the United States is falling behind other countries. In US schools, for example, only 41% of eighth-graders (13–14-year-olds) in 1999 had a maths teacher who had trained in maths, compared with an average of 71% internationally. Also, China produced more than 600,000 engineers last year, whereas the United States produced only 70,000.

► www.nationalacademies.org

UK science panel targets evidence for policies

Britain's key parliamentary science committee looks set to adopt a different strategy under its new leader, Liberal Democrat Phil Willis.

Willis, who took over on 20 July, says he wants the House of Commons Select Committee on Science and Technology to look harder at evidence used by the government when setting policy. The 11-strong panel will discuss possible topics next month, but Willis says he would like to investigate issues such as whether the

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Phil Willis has signalled a fresh course for the UK parliamentary science committee.

government has followed evidence on matters such as mobile-phone safety.

The committee has already pledged to study the technology that could be used to capture and store greenhouse-gas emissions from power stations.

Japan launches plans for supersonic jet travel

Japan took a small step towards developing a next-generation supersonic plane when an unmanned jet aircraft flew for 15 minutes at twice the speed of sound on 10 October over Woomera in Australia. The flight was powered by a rocket

booster (see *Nature* 437, 14; 2005).

Engineers sighed with relief as “everything went as planned”, says Keiichi Hirako, the experiment's leader at the Japan Aerospace Exploration Agency (JAXA). In a test three years ago, the plane separated from its rocket booster and crashed. JAXA aims to commercialize a supersonic jet capable of carrying 300 passengers by 2025.

Corrections

In the News story “Japan jumps towards personalized medicine” (*Nature* 437, 796; 2005), the description of aminoglycosides, drugs that can cause ear damage in people with a certain genetic mutation, was wrongly attributed to the cancer drug irinotecan. Both are candidates for Japan's efforts to create personalized medicine as described in the article. *Nature* wishes to apologize for the confusion.

Contrary to a statement in the Editorial “Responding to uncertainty” (*Nature* 437, 1; 2005), there is no evidence that lives have been lost as a result of the significant dip in take-up in Britain of the triple vaccine for measles, mumps and rubella.

Clarification

California-based Bio-Rad remains licensed to produce and sell thermal cycler products. The court injunction reported in “PCR spat broadens out” (*Nature* 437, 317; 2005) applies only to the sale of the products of its subsidiary, MJ Research, within the United States.

Life at the cutting edge

Biologists, planetary scientists and engineers have gathered in southern Spain to test a robotic drill. They hope some day to probe for life beneath the surface of Mars. **Jenny Hogan** investigates.

"O kay, we're at the surface," a voice calls across the tent. "Can someone hit the 'yes' button?" In the centre of a makeshift laboratory, sweltering beneath the Spanish sun, people cluster anxiously around a robotic drill. Its silver arm draws up from the deep hole it has made in the dirt. Hitting the 'yes' button will trigger the drill to release its cargo, and reveal whether it has managed to bore material from metres beneath our feet.

"Have we got a core?" asks Howard Cannon, an engineer from NASA's Ames Research Center in California. Peering into the end of the instrument, he spots a plug of red mud, some 7 centimetres long. This is good news. The drill is a prototype for an instrument that might one day be sent to Mars. It's Cannon's job to see if he can get the drill to work here on Earth.

A team of mission scientists is pretending that this drill, and the precious core it carries, is located on Mars. The drill is equipped with scientific instruments to search for signs of life in the dirt. So, for the team, extracting this chunk of red earth could be a step towards finding out whether life lurks beneath the sterile surface of Mars.

The drilling technologies being tested here, or others like it, may one day travel to the red planet aboard European Space Agency or NASA missions (see box, overleaf). Although the martian surface is too cold, dry and drenched in radiation for anything to survive above ground, past life forms could have taken refuge underground, some scientists think.

Data from the swarm of spacecraft currently roaming over and orbiting Mars have identified some possible spots where surface life may once have thrived. "We do our best to pick the sweet cherries, the prime candidates for life," says James Dohm, a planetary geologist at the University of Arizona, Tucson. "But eventually we have to have subsurface analysis."

The Spanish project is headed by planetary scientist Carol Stoker of the Ames Research Center, a veteran of previous Mars missions, and focuses on the unusual geochemical environment of Spain's Rio Tinto region.

To a visitor, the landscape looks little like barren Mars. The drill sits at the edge of an old



mining pit that is surrounded by pine trees. A lake of red-hued water lies below, and groves of orange trees stretch out in the distance. Local authorities recognize the beauty of the place; they erected a sturdy stone picnic bench for tourists — right on one of the team's preferred drilling sites.

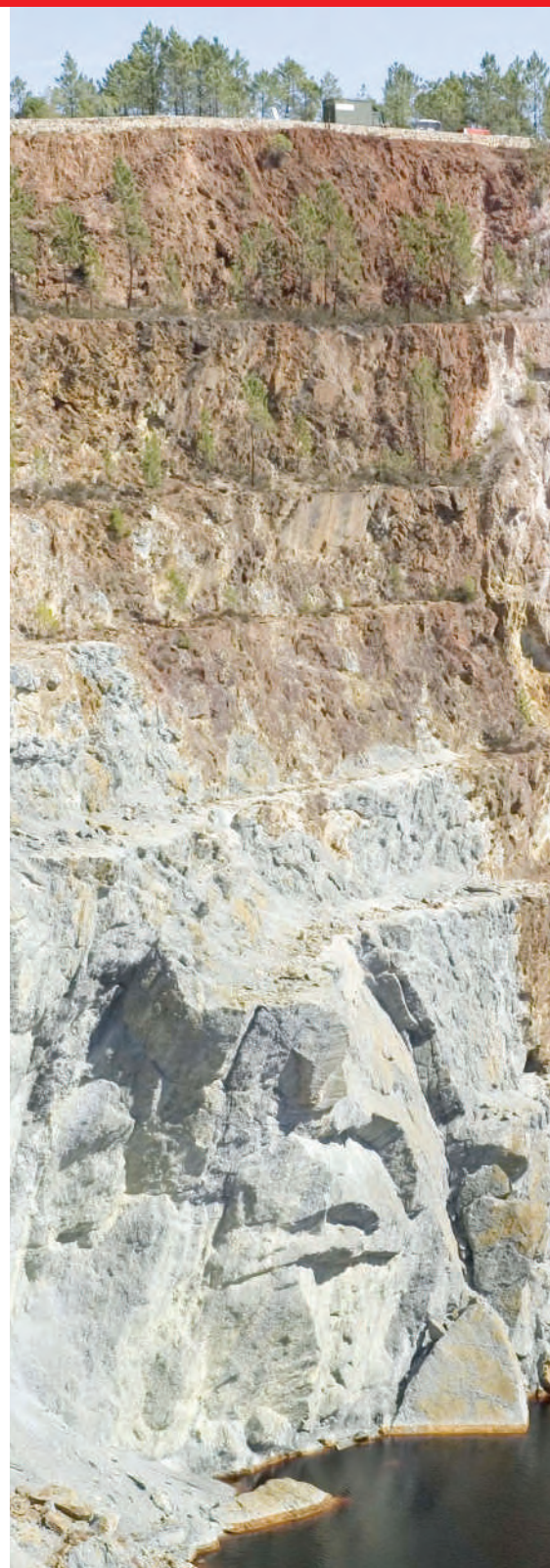
Hostile habitat

But it's underground, in the rocks from which the scenery is carved, that the Rio Tinto region may be most like Mars. A hardy microbial world inhabits the acidic waters of the Tinto river. This is what astrobiologists dream of finding in the subsurface of other worlds.

Microbiologist Ricardo Amils, of the Autonomous University of Madrid, a member of the science team, has studied the area for years and once gave Stoker an enthusiastic tour. Her imagination was fired by the idea of an underground ecosystem. "I'm not sure it struck him as a really good planetary analogue, but it did me," she says. Thus, the Mars Analog Research and Technology Experiment (MARTE) was born.

In 2004, the project's second year, it got a huge boost from data flowing back from NASA's Mars Exploration Rovers, Spirit and Opportunity. The rovers identified minerals on the surface of Mars, including the sulphate jarosite, which suggested that, when it existed, martian water was salty and acidic¹. These are just the kind of conditions found at Rio Tinto².

Planetary scientists, with their study subjects



so far out of reach, have long scoured Earth for otherworldly environments. For years, Mars researchers have been studying the life forms that eke out an existence in the Atacama Desert of Chile or in the cold, parched Dry Valleys of Antarctica. One group is exploring ice-bound communities of bacteria in a frozen volcano on Svalbard, the arctic archipelago off the coast of Norway. Some, in an effort to understand the martian atmosphere, are studying the dust devils whipped up in

E. JAMES/NASA AMES



A robotic drill that might one day be used on Mars is tested out at the edge of an old mining pit in Rio Tinto.

Nevada's Eldorado Valley. Others put rovers through their paces in rock-strewn deserts.

Michael Meyer, NASA's lead scientist for Mars, says it is important to study as many of these places as possible. "We don't want to get too far ahead of ourselves in predicting what Mars is like and expend all our energy on one site," he says. But Rio Tinto, he says, is a good place to include.

The Tinto river gets its unique chemistry from the Iberian Pyrite Belt, a vast deposit of



The analysis of cores extracted by a drill (left) may shed light on the nature of the scum in Rio Tinto's springs.

sulphide minerals tens of kilometres wide that runs for about 250 kilometres through southern Spain and Portugal. These minerals were laid down when the region was under water; at the site of a hydrothermal vent, deep-seated volcanism forced hot water, rich with dissolved chemicals, through cracks in Earth's crust.

Bottom dwellers

Although the Tinto river is naturally acidic and loaded with heavy metals, mining has exacerbated these characteristics³. For 5,000 years, locals mined the deposit for heavy metals, notably copper. Around Nerva, the Spanish town where the MARTE team is based, the hills are scarred by open pits and abandoned mining equipment. The acidity, in turn, led to other metals in the rocks dissolving, including the iron that gives the Tinto river its distinctive red colour and name.

Amils thinks that some of the river's heavy metals are extracted from their ores by bacteria living inside the rocks. This unique biosphere, he reckons, is powered by sulphur and iron. He says the river could serve as "an exhaust for a giant underground bioreactor". Although Amils and other biologists have quantified the microbial ecosystem at the surface^{4,5}, the MARTE project has been their first opportunity to go underground.

While Cannon is operating the robotic drill in the stuffy tent, Stoker and other members of the MARTE science team are enjoying the

air-conditioning of the ultra-modern Center for Astrobiology building outside Madrid. Olga Prieto Ballesteros, head of the remote-operations team, marks the new core on a whiteboard.

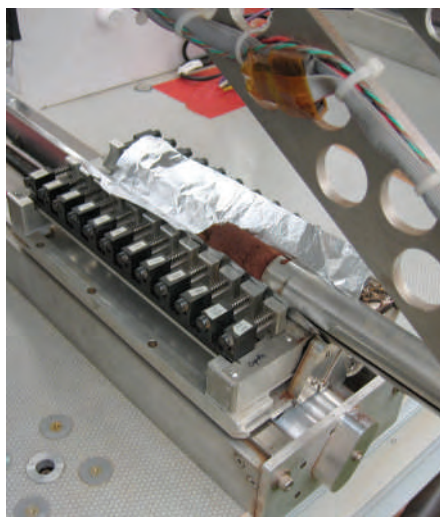
A satellite dish is relaying data from the drill site at a rate of 50 megabytes per day — roughly equivalent to the speed at which a spacecraft on Mars might be able to transmit information to Earth. The remote science team can access data from the drill's instruments, including pictures of the core and the output of an imaging spectrograph that helps identify the minerals present.

This year, the team's month of drilling nets them a hole just over 6 metres deep. Geologists may snicker to call this deep drilling, but Stoker calls it an impressive accomplishment for a robotic drill loaded with science instruments. "Nothing like this has been done before," she says.

If the remote researchers spot an interesting part of a core, they can get the robotic drill to subsample it, and pass the sample through sophisticated life-detection systems. The team can also instruct the drill to lower a camera into the borehole, to inspect the surrounding walls. In this case, the team knows there's a good chance that life could exist in the borehole. But it pretends otherwise. "If we see a filament in the rock, we know it's a root," says Ballesteros. But the root is sampled and tested before they conclude that it is life.

Of the samples analysed already, "we have detected life in cores three and five", says a team biologist. The results come from the signs-of-life detector (SOLID) instrument, which uses a protein microarray to look for molecules that are characteristic of life. Astrobiologists routinely cite such techniques

"The MARTE project has advanced the date the technology might be ready for a drilling mission on Mars." — Carol Stoker



Carol Stoker (left) hopes that the drill technology developed by her team could be adapted for Mars.

when they talk about searching for life on other planets⁶.

Back at Rio Tinto, the drill is working hard to bring up new samples. It is designed to run with minimum human intervention. "Right now, we're sitting here doing nothing," says Cannon. "But every once in a while something will fail and we have to step in."

Terrestrial drillers would usually select their drill bit to suit the rock, and change the speed and power depending on the conditions they encounter. In contrast, a robotic drill on Mars is stuck on its own with just one drill bit and one technique.

Another huge challenge is cleaning out the hole, to prevent the drill from clogging. On Earth some kind of fluid can be pumped down the drill shaft to remove the cuttings. This isn't feasible on Mars, so the drill built for

the MARTE by New York-based Honeybee Robotics operates dry.

Lone worker

A screw around the outside of the drill bit is meant to carry the debris to the surface, but it keeps getting encrusted with muck. The team has resorted to vacuuming it clean. Red dust coats everything, including the laptops running the robot.

As the robot ejects a new core, a researcher lunges inside the rig to catch it on sterilized tin foil. Moments later, she dabs a cotton bud over the core, swabbing material to analyse for ATP, the energy-rich molecule that acts as fuel for life on Earth. Samples destined for the SOLID instrument are rushed from the drilling tent to an air-conditioned vehicle next door, because the tent is too hot

for the instrument to operate properly.

Despite such difficulties, the MARTE team remains optimistic. They had not set out to solve every drilling and scientific question in 30 days. The point of the simulation, they say, was to see what works and what needs to be worked on. "We've got a lot of work to do, but this is a first step," says Cannon. Stoker agrees, and is enthusiastic about what they have learned: "This project has advanced the date the technology might be ready for a drilling mission — there's no doubt about that." Still, the project's US\$6-million funding runs out this year, and it's unclear whether the drilling rig will get another outing.

Biologists will be back at the Rio Tinto site, regardless. They may have drilled deep holes, but Amils says they've only scratched the surface of understanding how this ecosystem works. During a tour of some springs near the source of the Tinto, he notes the murky red, brown and white deposits in the water and on the surrounding rocks. Amils labels many of the oozy, lumpy patches 'scum', a euphemism for 'difficult to identify'. Some are minerals, some are life. Most are not yet classified.

Todd Stevens, his colleague and an expert in geomicrobiology from Portland State University in Oregon, sums it up: "We have a lot of work to do to understand the interactions of the rocks and microbes on Earth before we can do it on other planets."

Jenny Hogan is a reporter with *Nature*.

1. Klingelhöfer, G. *et al. Science* **306**, 1740–1745 (2005).
2. Fernández-Remolar, D. *et al. Planet. Space Sci.* **52**, 239–248 (2004).
3. Davis, R. A. *et al. Environ. Geol.* **39**, 1107–1116 (2000).
4. González-Toril, E. *et al. Appl. Environ. Microbiol.* **69**, 4853–4865 (2003).
5. Amiral Zettler, L. A. *et al. Nature* **417**, 137 (2002).
6. Whitfield, J. *Nature* **430**, 288–290 (2004).

FIRST DRILL ON MARS

The first drill to reach Mars might be launched in 2011 by the European Space Agency (ESA), on a mission called ExoMars.

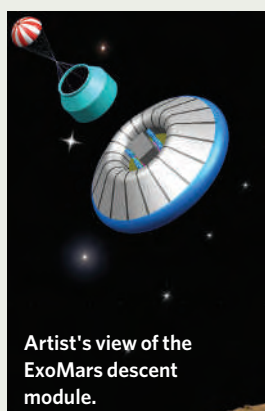
If approved at a December minister-level meeting, ExoMars would include a small rover carrying a lightweight drill. It would be capable of boring 2 metres into the martian surface. That would be deep enough, scientists hope, for it to get through the oxidized surface layer and into relatively unaltered martian soils.

It's certainly deeper than any sampling tool has ever gone before. The Viking landers scratched up some topsoil in the 1970s, and today's Mars rovers grind small holes in surface rocks. Only the Beagle probe, presumed to have crashed on landing in 2003, carried the

technology to get underground, and its 'mole' never got a chance to burrow.

ESA already has experience in putting drills aboard spacecraft. Last March, it launched the cometary mission Rosetta, which carries a drill to pierce more than 20 centimetres into Comet Churyumov-Gerasimenko.

For the ExoMars project, ESA is considering two possible designs. One, named DeeDri, is being developed by the group that provided Rosetta's drill,



Artist's view of the ExoMars descent module.

headed by the Italian Space Agency. The other, CanaDrill, is being sponsored by the Canadian Space Agency.

CanaDrill could have the first shot at getting into space, if it is selected for NASA's unmanned lunar lander, details of

which were announced last month. The mission could launch as early as 2010, to search for water in the shadowy craters near the Moon's south pole.

With NASA's new push to send astronauts to the Moon

and Mars, the agency is more interested than ever in nurturing drilling technologies. Drills could help astronauts reach natural resources, such as underground water reservoirs if they exist.

The challenge is making drills lightweight and energy-efficient enough to fly, says David Beaty, a former exploration geologist in the oil industry who now works for NASA's Jet Propulsion Laboratory in Pasadena, California. He is organizing a 'drill-off' that will take place next February, when a handful of companies developing robotic drills will test their machines at Idaho Falls.

Competitors will be charged to build a drill that weighs no more than 40 kilograms, digs 20 metres deep, and runs off the power of a lightbulb.

ESA



Patchwork people

For years it was assumed that tiny differences in our genetic make-up gave us our individual traits. Now it seems that those characteristics are caused by rearrangements of large chunks of our DNA — variations that could be the key to understanding disease. **Erika Check** investigates.

Exactly one year ago this week, scientists announced that they had finished the 'Book of Life'. The complete sequence of the human genome had been painstakingly reduced to an ordered list of letters representing the four bases of DNA. This text was believed to be virtually identical for every person on Earth — and the major differences between individuals, such as hair colour, were said to be the equivalent of typographical errors, no longer than a single letter. The next major task for scientists was to find out which of these tiny differences can cause disease.

But even as the ink was drying on the complete sequence, some researchers were ques-

tioning whether there was really such a thing as the definitive edition of the Book of Life. By skim-reading individual genomes, these scientists were finding bizarre and unexpected irregularities. In some people, whole paragraphs of the text were duplicated, whereas in others, large passages were missing, or even printed backwards. These major revisions turned up in all kinds of people, including many who seemed healthy and normal. Suddenly, it seemed possible that there was actually no standard version of the Book of Life, and researchers wondered whether we are all much more different from each other than they had thought.

Now, scientists are beginning to link these larger variations in the genome to human health and disease. Some even say that the variations explain normal human diversity and evolution, which have always seemed too intricate to be the result of differences in just single DNA bases. "When we first looked at the human genome, we were so very proud of being able to look at the definitive sequence," says genomics researcher Chris Ponting of the University of Oxford, UK. "Now, in just a few years, we've travelled such a long way. We've gone from looking at the human genome to looking at human genomes, plural."

The widespread existence of all these varia-

tions is a big surprise because such large changes have been associated with devastating genetic diseases¹. Deleting a chunk of DNA, for example, can eliminate important genes. Extra copies of a gene can cause overproduction of a protein, throwing the cell's finely balanced biochemistry out of kilter. And moving a DNA chunk from one location to another, or even reversing its orientation, can dramatically change the signals that control genes.

So far, most of these variations have not been found to cause overt disease. But researchers now think that they may have a more subtle role to play. They could influence our vulnerability to diseases that are caused by a complex mix of genes and environment. To find out more, two major public projects are cataloguing the extent of these variations in normal humans. One, funded by the US National Institutes of Health, will look for structural variation in the genomes of ten individuals from different ethnic backgrounds. The other, funded by Genome Canada and the UK-based Wellcome Trust, will scan the genome for genetic differences among the 270 people included in the International HapMap Project — a huge study looking at human genetic diversity. And several private donors are funding projects to find out whether the large variations might cause complex diseases such as cardiovascular problems, Parkinson's, psychiatric disorders and autism.

Shock results

The pioneering work that spurred these large projects began in 2002. Data were pouring out of the efforts to sequence the human genome, as were tools that allowed scientists to compare DNA from different individuals in incredibly fine detail. At first, researchers were surprised and even disturbed by the new findings.

Charles Lee, a cytogeneticist at Brigham and Women's Hospital in Boston, Massachusetts, was investigating whether he could use one of the new technologies as a genetic test. But his experiments kept failing. He frequently found major aberrations in the gene sequences of normal patients he was trying to use in the control group. Some of this group were apparently carrying more copies of certain genes than others, yet they seemed perfectly healthy.

Lee was unsure about what was going on until, in late 2003, he went to Canada to give a talk at a meeting in Toronto. There he discovered that Steve Scherer of the Hospital for Sick Children in Toronto was seeing the same weird phenomenon: normal, healthy patients with different copy numbers for certain genes.

Meanwhile, at Cold Spring Harbor Laboratory in New York, molecular geneticist Michael Wigler was using a different technology to compare the genomes of two men: a Caucasian and an African pygmy. He also saw copy-number changes where he did not expect them — in this case, in a gene that codes for a crucial brain chemical. "We were very excited, and then very afraid," says Wigler,

who was worried that he had discovered a mutation that would predispose one of the men to schizophrenia.

But in the next experiment, his group found copy-number changes in a different gene that is active in human sperm. "That really freaked us out," Wigler says. Over the next year, his group turned up more examples of the same phenomenon. "It didn't take us long to realize we were looking at something that was fairly common," Wigler says.

How common, exactly? Last July, Wigler's group reported that it had looked at 20 normal individuals and found 221 places in the genome where those people had different copy numbers of stretches of DNA². Some of these copy-number changes showed up in more than one person, and so qualify as

'polymorphisms' — shorthand for particular spots in the genome that regularly differ between individuals. In the Book of Life analogy, these polymorphisms represent sections of text where certain paragraphs are repeated different numbers of times in different individuals.

About 76 of the variations Wigler's team found were polymorphisms, and each person had about 11 of them in his or her genome². Soon after, Lee and Scherer reported that in a survey of 55 people they had found 255 copy-number variants, 102 of which were polymorphisms³.

Different strokes

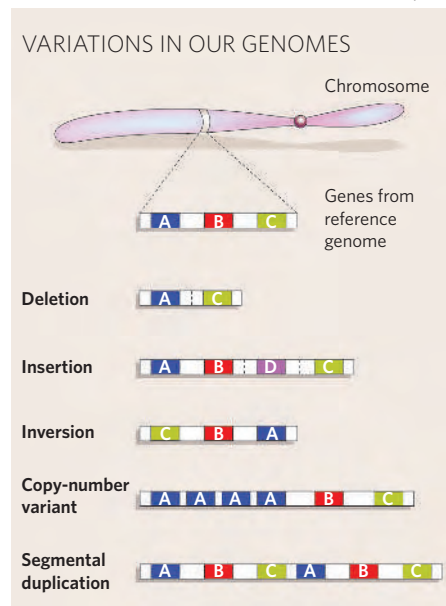
But copy-number polymorphisms tell only part of the story. Earlier this year, a team led by Evan Eichler at the University of Washington in Seattle published evidence of rampant changes of a different sort⁴. Eichler's team compared a portion of an individual woman's genome with the 'reference' human genome sequence produced by the Human Genome Project. The team found 297 potential places where DNA had been rearranged in one of the genomes. Some of these rearrangements were insertions or deletions of stretches of DNA. Others were inversions, where a long stretch of sequence had been reversed in one of the genomes.

To go back to the Book of Life, these places represent missing or repeated pages, or whole sections of text that read backwards (see graphic, left). "We can see this whole new world of variation," Eichler says. "It's going to give us a fresh view of the comprehensive genome, from the single base-pair differences to the really large variants."

Genome researchers now have a catch-all

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Biology lesson: our ability to learn may be governed by large differences between our respective genomes.



phrase for the vast array of rearrangements — including copy-number polymorphisms, inversions, deletions and duplications — that occur normally in the human genome. They call it structural variation, and have described at least 800 individual variants that, in total, account for about 3.5% of the human genome. And the sheer number of variants seems likely to catch up with the number of known single nucleotide polymorphisms — the single-letter ‘typos’ in the Book of Life. That makes structural variation a potentially major source of diversity. It is even possible that we’re not all 99.9% similar, as the Human Genome Project predicted.

The biggest question about structural variation is: does it matter? There are already some hints that it does. Eichler’s analysis this year showed that many of the genes found in structural variants negotiate our interactions with the environment⁴. Some make proteins that break down drugs, for example, or help our immune systems respond to disease. So it makes sense that some of these variations explain our unique responses to the stresses or pleasures of life.

Strength in numbers

In March, a team headed by Sunil Ahuja at the University of Texas Health Science Center, San Antonio, published a new and dramatic proof of that principle⁵. The researchers analysed the average number of copies of an immune-system gene in African, European, Asian and American populations. They found that extra copies of the gene, which makes an immune-system protein called CCL3L1, helped protect people against HIV. If patients with the extra copies became infected by HIV, they progressed more slowly towards full-blown AIDS than those with fewer copies.

There is also good reason to hope that structural variation will shed new light on complex diseases, such as obesity, whose development is triggered by the interaction of many genes, rather than one or two. For one thing, studies such as Eichler’s show that structural variation is involved in our response to the environment — a key factor in complex disease. What’s more, statistical analyses show that regions of structural variation contain genes that are still evolving in humans⁶. If these genes are important enough for evolution to be changing them, they must affect us in some way, for better or worse. And if we can’t detect these effects immediately, it is possible that regions of structural variation are interacting with other parts of our genomes in subtle ways to influence our most crucial traits.

“Given that everywhere we’ve looked hard, we’ve found that copy-number variation influences human disease, it would be strange if complex diseases didn’t also appear on that radar,” says evolutionary geneticist Matt Hurles at the Wellcome Trust Sanger Institute near Cambridge, UK. It’s this hope that is propelling the disease studies already under



IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Structural variation could help explain why some people are prone to obesity (top) and may lead to fresh therapies for this and other diseases.

way — most notably, Cold Spring Harbor’s \$11-million hunt for copy-number polymorphisms related to autism.

But the link between structural variation and natural selection raises perhaps an even more intriguing possibility. Has structural variation created new mutations that have helped to shape human evolution?

Brought to book

Eichler has found that copy-number variants often occur in larger blocks of repetitive DNA called segmental duplications⁸. These blocks take up about 5% of the human genome, and occur in the same places in all of our genomes. They also seem to cause trouble. Scientists believe segmental duplications make it difficult for our genome to replicate itself faithfully during the processes that create eggs and sperm. Mistakes result in DNA rearrangements, such as deletions and inversions, meaning that new structural variants are created and passed down the generations.

Icelandic scientists reported earlier this

year, for instance, that 20% of Europeans carry a large genetic inversion that is spreading throughout the population⁹. Women who carry the inversion have more children than those who don’t — a classic sign that it confers some sort of selective advantage.

And studies comparing us with our chimp cousins have already linked structural variation to our divergence from the apes. Last year, scientists from the University of Colorado in Denver and Stanford University found 1,005 genes that differed in copy number among humans and four other primates¹⁰. This month, Eichler’s group reported 651 likely structural rearrangements between chimps and humans¹¹. The group counted 245 genes contained in these variants, including some genes involved in reproduction and drug metabolism. Eichler’s group has also found that segmental duplications have created much more of our genomic differences from chimps than single base-pair differences⁷. There are 177 genes contained within the human-specific duplications. As such duplications are hotspots for evolution, those 177 genes could be partly responsible for creating the traits that make us human.

These genetic differences could also be useful. Scherer’s lab has just released a targeted analysis of inversions between the chimp and human genomes¹². The group found 1,576 probable inversions, and confirmed 23; three of these differed among human individuals. Not only does this shed some light on primate evolution, but as inversions can often predispose DNA to harmful mutations, these inversions might be involved with human disease. “If you can highlight the structural variations that are inversions, you might be able to highlight where you should look for regions involved in disease,” Scherer says.

In other words, those quirky differences written into the pages of the Book of Life are more than just nonsensical scribbles. Those weirdly rearranged sentences may actually portend life or death. Deciphering the meaning of these cryptic passages could help scientists to diagnose, prevent and treat disease — although that will probably take many years. For now, the realization that we are all reading from individual texts has already altered scientists’ understanding of humanity — and of the library of unique volumes that makes up the human race. ■

Erika Check is Nature’s Washington biomedical correspondent.

“This whole new world of variation is going to give us a fresh view of the comprehensive genome.”
— Evan Eichler

1. Buckland, P. R. *Ann. Med.* **35**, 308–315 (2003).
2. Sebat, J. et al. *Science* **305**, 525–528 (2004).
3. Iafrate, A. J. et al. *Nature Genet.* **36**, 949–951 (2004).
4. Tuzun, E. et al. *Nature Genet.* **37**, 727–732 (2005).
5. Gonzalez, E. et al. *Science* **307**, 1434–1440 (2005).
6. Check, E. *Nature* **435**, 252–253 (2005).
7. Cheng, Z. et al. *Nature* **437**, 88–93 (2005).
8. Sharp, A. J. et al. *Am. J. Hum. Genet.* **77**, 78–88 (2005).
9. Stefansson, H. et al. *Nature Genet.* **37**, 129–137 (2005).
10. Fortna, A. et al. *PLoS Biol.* **2**, E207 (2004).
11. Newman, T. L. et al. *Genome Res.* **15**, 1344–1356 (2005).
12. Feuk, L. et al. *PLoS Genet.* preprint No. 2005, e56 doi:10.1371/journal.pgen.0010056.eor (2005).

BUSINESS

Venture capitalists tackle Chinese hurdles

Research campuses in Beijing and Shanghai are brimming with ideas that might invite commercial development. And for the first time, venture capital to take them forward is pouring into the country.

Last year, according to Ernst & Young, venture-capital investments in China totalled US\$1.3 billion — more than those in France, if only a fraction of the \$20 billion raised in the United States. Yet according to a report released last month by the international consultancy, the Chinese venture-capital market remains fragile.

“Chinese venture capital is at a crossroads,” the report concludes. “Unless the conditions necessary to sustain a healthy venture-capital industry are put in place, the pace of investment is likely to slow or even be reversed.” Indeed, the report suggests, an 8% drop in venture-capital investment during the first half of this year may be an ominous sign.

Venture capital in China first got going in 1992, when IDG Technology Venture Investment began investing in information-technology companies there. The field has grown rapidly of late, trebling in value since 2002. Yet only a small fraction of this is going into science-based businesses. Many of the ventures backed are conventional services and manufacturing ones, says Yoshiaki Hasegawa of the Tokyo-based Japan Asia Investment Company. And of the technology firms backed by venture capitalists, most are in semiconductors and information technology.

The report — written by Ernst & Young’s Gil Forer on the basis of a Shanghai workshop co-hosted with the China Venture Capital Association — warns that investor interest is being dampened by venture capitalists’ inability to make money from small Chinese companies by floating them on a stock market. It calls on the Chinese government to set up a stock market similar to the Nasdaq. But earlier this year, the government strengthened restrictions on Chinese companies raising money through share offerings on foreign stock markets. “It is trying to control the flow of capital,” says Forer.

International investment funds such as IDG and the SoftBank Asia Infrastructure Fund, based in Hong Kong, account for about 70% of total venture-capital investment in China, according to Ernst & Young. The majority of the 200 domestic venture-capital funds that

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

At the crossroads: venture capitalists are getting mixed signals from the Chinese government.

account for the rest are closely tied to city and provincial governments. And foreign funds tend to follow their lead. “They want assurance” concerning the reliability of target firms, says Ming-Wei Wang, president of SiniWest, a biotechnology company based in San Diego with interests in China and the United States.

According to Wang, government investment is often spread too thin, as it tries to please too many people. And government-associated funds often base their investments on personal connections, rather than potential profitability. “We need independent evaluation,” he says.

Researchers themselves can put off venture capitalists by refusing to share information, Wang adds. Such refusals are understandable, given the weak legal mechanisms that are available to protect intellectual property in China, but they make ideas hard to evaluate.

Guo-Ping Zhao, executive director of the Chinese National Human Genome Research Center in Shanghai, says that venture funds in China often lack the know-how to evaluate their investments. “Those who have the money don’t understand technology,” he says. “Those who understand it don’t have any money.”

There are some signs that the government would like to foster venture capital: last month, it actually eased some restrictions on citizens’ activities on foreign stock markets. And seasoned investors are confident that the market will continue to expand. “China is moving in a positive direction,” says Hasegawa, whose firm has already invested in two dozen Chinese companies. ■

David Cyranoski

M. HENLEY/PANOS

IN BRIEF

PATENT PEACE Brazil and Abbott Laboratories have averted a patent showdown by announcing that they have agreed on pricing for a key AIDS drug.

Beginning next March, the Illinois-based drug company will make the protease inhibitor Kaletra (lopinavir and ritonavir) available to the Brazilian government for 63 US cents a pill, rather than the current price of \$1.17.

In June, Brazil’s health ministry threatened to break Abbott’s patent — which doesn’t expire until 2015 — and begin producing a generic version of the drug if the company didn’t lower its price. Kaletra is a staple in Brazil’s renowned, publicly funded AIDS programme, which distributes free drugs to roughly 160,000 patients.

GOING OFFSHORE Industrial research and development is rapidly going global, according to a survey by the Paris-based Organisation for Economic Co-operation and Development.

The survey says that 16% of industrial research is now performed by overseas affiliates; the concentration of such ‘offshore’ research is highest in Ireland and Hungary, and lowest in Japan.

The economic think-tank says that of the largest economies Britain has the most internationalized industrial-research system by three separate criteria. China has become the third-largest research nation in the world, and surpassed Japan as the country with the second-largest research workforce.

CANCER CASH California biotechnology company Genentech cruised to a record-breaking third quarter on booming sales of anti-cancer drugs. Its net income increased by 56% from the third quarter in 2004, and for the first time, US sales of the company’s cancer drugs exceeded \$1 billion.

The strong growth was led by US sales of Avastin (bevacizumab), a colon-cancer drug that starves cells of their blood supply. US sales of Avastin grew by four-fifths, to \$325 million, compared with the same time period in 2004. Sales of Herceptin (trastuzumab), a genetically targeted breast-cancer drug, grew by 70% to \$215 million.

Don't underestimate the death rate from Chernobyl

SIR — Your News story “Chernobyl: poverty and stress pose ‘bigger threat’ than radiation” (*Nature* 437, 181; 2005) suggests that the health and environmental effects of the Chernobyl accident were not as great as originally suggested.

Writing on behalf of an international group of researchers in this area (see <http://cricket.biol.sc.edu/chernobyl/nature/letter.pdf>), we believe that these suggestions, based on the reports of the UN Chernobyl Forum, are misleading.

“As we approach the twentieth anniversary of the Chernobyl disaster we should be sensitive to the long-term implications.”

— T. A. Mousseau and colleagues

The full estimate, given by the UN report, of people who could eventually die of factors linked to radiation includes people in other contaminated areas as well as those within Soviet Contaminated Zones and is 9,335, not 4,000 as reported. This estimate is similar to earlier estimates of future cancer mortality prepared by the US government in December 1987 (*Report on the Accident at the Chernobyl Nuclear Power Station*, US Nuclear Regulatory Commission, Washington DC). Further details to support our argument that neither of these estimates should be downplayed are available at the website above.

We believe it is too early to assess the overall impacts of radionuclide exposure on human health or on plant and animal populations. In particular, we do not know all the possible consequences of the multi-generational accumulation of genetic defects. As we approach the twentieth anniversary of the Chernobyl disaster we should be more sensitive to the long-term implications rather than suggesting that the coast is clear for redevelopment in the contaminated zones.

Up to now, most studies have focused on cancer, because of funding constraints, with little investment in studies of non-cancer morbidity or model systems. But model organisms with relatively short lifespans may provide a clear picture of the multi-generational consequences for human health, while humans exposed to Chernobyl are a unique population that must be supported and observed far into the future.

Given the long latency period for many diseases and the growing interest in rejuvenating the nuclear power industry, it is imperative that studies of the affected populations continue.

Timothy A. Mousseau*, Neal Nelson†, V. Shestopalov‡

*School of the Environment,

University of South Carolina, Columbia, South Carolina 29208, USA
†8102 Ashton Birch Drive, Springfield, Virginia 22152, USA

‡Radioecology Centre of the Ukrainian National Academy of Sciences, Kyiv 01054, Ukraine

Media reports: call for a working party is unrealistic

SIR — The call, made in your Editorial “Responding to uncertainty” (*Nature* 437, 1; 2005), for learned societies to create fast-acting working parties to respond to sensational news stories arising from journals, is unrealistic in the United Kingdom.

Having worked in and with UK biological learned societies for a couple of decades, including a term as a member of the Council for the Association of Learned and Professional Society Publishers, I can testify that the vast majority simply do not have the resources to implement this suggestion. Further, their members have full-time commitments and so would find it difficult to attend such working parties at short notice.

The best that might be achieved would be for those societies that publish journals to call upon their editorial boards to justify the decision to publish, and to clarify any uncertainty arising out of media reporting. However, even this would not be applicable to all cases: for instance, the example you cite of the MMR vaccine would not be covered as *The Lancet* is produced by a publishing house and not a learned society.

Jonathan Cowie

Science-com.concatenation.org,
Northumberland Heath,
Kent DA8 3EU, UK

Main effect of bureaucracy is to reduce productivity

SIR — Your News Feature “The nightmare before funding” (*Nature* 437, 308–311; 2005), on the horrors of writing grant applications, will strike a chord with researchers everywhere.

To my knowledge there has never been any objective analysis to determine whether or not increasing the complexity in the funding process has improved decision-making by the funding agencies or produced higher-quality research from the recipients of the funds. The only outcome of which we can be certain is that the wasted time makes grant-writers less productive. This in turn means that, for those agencies that burden us with this administrative nonsense, the science base is less competitive. I know this because,

in order to keep my own research group running on a relatively modest scale, I have not been able to do an experiment for five years, and I am certainly not alone.

The pessimistic view is that we are now too bogged down in this mess ever to return to a simpler, more sensible way of doing things. I tell my PhD students that the allure of research used to be the intellectual challenge, but nowadays you'd better be up for the administrative challenge too.

Stephen Moss

Ashton Chair of Biomedical Research,
Division of Cell Biology,
Institute of Ophthalmology,
University College London,
11–43 Bath Street, London EC1V 9EL, UK

Concern at animal research should not be dismissed

SIR — Your Editorial “Still not deterred” (*Nature* 437, 1–2; 2005) calls on universities to support animal research. We are all horrified at the techniques of the terrorists, but the much wider public is increasingly sceptical about animal research. These are legitimate concerns that should not be dismissed because of a few fringe characters. As scientists, we depend on public funds, so we are under an obligation to explain why animal research is morally acceptable.

We argue that animal research can yield information not available through other experimental techniques. Historically, we have experimented on slaves and prisoners. Given that we can no longer justify the suffering of human research subjects, even if it could save lives, how can we accept the suffering of animal research subjects? We must recognize that the decision to experiment with animals is a moral decision, not a scientific one. Our society does not believe that the end justifies the means. Do we disagree?

Our contention that humans are unique among the animals is a religious argument. As an evolutionist, I find it absurd to think that human culture and society developed without precedents among animals. We in fact now have evidence for animal cultures, as reported in the pages of this journal.

Perhaps we need to come up with better arguments — or better experiments that don't offend the public.

Jacqueline Zupp

CCBC - F520, 7201 Rossville Boulevard,
Baltimore, Maryland 21237, USA

Contributions to Correspondence may be submitted to corres@nature.com. They should be no longer than 500 words, and ideally shorter. They should be signed by no more than three authors; preferably by one. Published contributions are edited.

BOOKS & ARTS

Oppenheimer: the opera

A musical treatment of the development of the atomic bomb raises questions still relevant today.

Doctor Atomic

composed by John Adams, directed by Peter Sellars

San Francisco Opera world premiere, 1 October 2005

www.doctor-atomic.com

Philip Campbell

For an opera to succeed, the music not only has to have a stature of its own, but also needs to supply a deeper insight that could not be expressed by a spoken narrative. In other words, it must seem truly necessary. Challenging under any circumstances, this goal for a composer is all the more so when the opera's purpose is to reflect the states of mind of the people who created the first atomic bomb, in the weeks and hours leading up to the 'Trinity' test in the Alamogordo desert.

John Adams' opera *Doctor Atomic*, which was premiered earlier this month, seemed to my ears ultimately to succeed in that goal. The audience in the San Francisco Opera house certainly gave it enthusiastic acclaim. The soloists, chorus, dancers and orchestra, under the strong leadership of director Peter Sellars and conductor Donald Runnacles, did the piece proud. Readers should see it if they get the chance.

There are four narrative elements at the core of the opera: the moral concerns of the physicists, realizing that Germany had been defeated and facing the prospect of the bomb being used on Japan; the preparations and tensions in the small community charged with delivering a working device, now needed ahead of President Harry Truman's discussions with Joseph Stalin and Winston Churchill; the emotional states of Robert Oppenheimer — the technical leader of the project — and those close to him, especially his wife Kitty and children's maid, Pasqualita; and an extended passage at the end when time stretches out, in the moments leading up to the explosion.

Each of these elements is conveyed by distinct means in the opera. It successfully conveys the industrial scale of activity in the preparation of the device through (for example) scurrying descending scales in the orchestra accompanying choral chants on the principles and power of the weapon. This is all visually amplified on the stage by movements of the chorus and dancers. Here, Adams' repetitive patterned style of writing is perfectly suited



The spectre of the bomb hangs over Oppenheimer and those around him.

to the task, as it is also in the opera's climactic representation of the endless moment just before the detonation.

To express the charged emotional states of Oppenheimer and his household, Sellars, the opera's librettist, used poetry — extracts from Charles Baudelaire, John Donne, the *Bhagavad Gita* and the US poet Muriel Rukeyser. This approach has some justification, as poetry was second nature to the Oppenheimers. These texts draw from Adams the most beautiful music in the opera, at times orchestrated with a French feel for sensuality and delicacy.

But there is a downside. The texts are elusive in their meaning — they require an extra layer of interpretation from a listener already faced with the challenge of following an operatic narrative. This task would be impossible without having the texts displayed in surtitles, as they were in San Francisco. Even so, it is difficult to be drawn into the emotions with immediacy, despite the lyrical strength

of Adams' music. A similar distancing arises at other instances as a result of the music. While busy on its surface, the music is occasionally little more than a slow succession of static chords embellished by the singers' matter-of-fact sequences of notes. Lyricism is not essential in opera, but there seems occasionally to be a need for a more vivid integration of the musical elements. Unlike some contemporary composers, Adams builds on the expressive power of sequences of chords, but here the compelling power slackens.

The musical language of the opera is direct, and as a whole it has a strong structure with reference points of recapitulation that relate to the state of the action, and a series of scenes of suitably contrasting character. The opera never seeks crassly to judge the character of Oppenheimer (Doctor Atomic), who is portrayed as personally sensitive yet consistently discourages his scientists from communicating their concerns: he remains an enigma. As Sellars explained in an eloquent introductory lecture, the opera steps up to the complexities of both Oppenheimer and the issues of science's power. As such, it provides a welcome contrast to the superficiality, as he sees it, of contemporary politics and culture.

Does the opera have a timeliness beyond the fact that we have reached the 60th anniversary of these events without being annihilated? Yes, on several levels. Adams characterizes this and his other operas (*Nixon in China*, *The Death of Klinghoffer*) as symbolic. The subject of the atomic bomb, the meeting between Nixon and Mao Zedong, and the events on the *Achille Lauro* appeal to him as an archetype, he says,

T. MCCARTHY

“summoning up, in choice symbols, the collective psyche of our time”.

But the opera is also pertinent to issues that often arise in *Nature's* pages. The application of science is ever more extended into the workings of the world and is all the more capable of instilling distrust in the society that pays for it. We should expect non-scientists to be wary of a profession that is often as excited, driven and single-mindedly focused as the Manhattan Project scientists.

The end of the opera should leave receptive audiences more aware of the case for scientists engaging with the public. The opera is framed at the beginning and end of each act by electronic soundscapes. The last moments take us beyond the Trinity test to the voice of a Japanese woman begging for water. This is not only a fitting reference to the ethical question that

overshadows the opera, but also a reflection of the societal challenge facing any scientist or technologist today whose research could change people's lives. Scientists still face equivalent moral debates of expediency (is the work justified by the benefits?) and of fundamental principle (is it in opposition to humanity's deepest values and hence to be forbidden?). They can choose to lock themselves away or consider the full implications of what is being achieved and reach out.

The efforts of leading scientists such as Robert Wilson and Leo Szilard to engage with decision-makers over Hiroshima, however debatable, should inspire us today. Adams' opera has helped them to do so in a way that no conventional history can. His music really is necessary.

Philip Campbell is editor-in-chief of *Nature*.

Mathematical musings

Euclid in the Rainforest: Discovering Universal Truth in Logic and Math

by Joseph Mazur

Pi Press: 2005. 352 pp. \$14.95, £10.99

Timothy Gowers

Euclid in the Rainforest by mathematician Joseph Mazur is difficult to summarize, as its title perhaps suggests, but here is an attempt. Mazur's central theme is reasoning, both mathematical and scientific, and he divides it into three kinds. First he discusses simple syllogistic reasoning of the kind made systematic by Aristotle, of which the most famous example is the following deduction: all men are mortal; Socrates is a man; therefore, Socrates is mortal.

Next, he shows that this sort of logic is not powerful enough to deal with an essential part of mathematics: the infinite. For thousands of years, mathematicians had to rely on intuition and common sense when, for instance, they

were discussing calculus or sums of infinite series. It was not until the late nineteenth and early twentieth century that formal systems were developed that provided the necessary rigorous underpinning.

The standards of proof in science are very different from those in mathematics, and in the third part of the book Mazur discusses less formal reasoning: what it is that makes certain statements highly plausible, even when it is inappropriate to look for formal proofs. For example, why is it rational to believe that a political party that is well ahead in the opinion polls just before an election will probably win it? Or why am I confident that when I go upstairs I will not fall through the floor?

What this summary of the book fails to convey is any sense of how Mazur writes about these ideas. He does not just dive in and try to explain them as clearly as possible. Rather, he tells stories, many of them about his own life,

which seems to have been very interesting — from time to time one wonders whether he has partly fictionalized it. He then lets his preoccupations emerge, much as a novelist might.

If you are interested in different forms of reasoning and are looking for a systematic treatment, this book is not the place to find it. For instance, bayesian analysis has a lot to say about plausible reasoning, but Thomas Bayes is mentioned only once, very briefly, in the introduction. Instead, this book is more like a fascinating conversation that stays within certain bounds but nevertheless often moves in unexpected directions.

For example, Mazur discusses, and clearly cares deeply about, a technique called 'facilitative communication', which allegedly allows severely autistic children to communicate with the help of a trained intermediary. Sceptics, of whom Mazur is one, ask why we should believe that messages that result from the technique come from the child, rather than the intermediary — a question of some importance, because evidence based solely on facilitative communication has been used to send parents to jail for sexual abuse. The reader is not offered a general method for deciding whether or not a statement is plausible, but the examples that bubble up from the narrative are thought-provoking in many ways.

These qualities make *Euclid in the Rainforest* an unusual book. It even contains, briefly and tastefully, a sex scene, which is probably a first for a popular mathematics book. There are also a few spelling mistakes (for instance, the playwright David Auburn, the philosopher Carl Hempel and Blecker Street are all misspelled), but these contribute to the conversational feel of the book — as if it's the sound that really matters — and add to its charm. Rather than trying to extract general conclusions from the book, the best approach is simply to relax and enjoy the company of its author.

Timothy Gowers is at the Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge CB3 0WB, UK.

Up close and colourful

Coral reefs are fascinating places. Marine biologists are still seeking to understand the complexity of life within these ecosystems, which also provide a valuable physical barrier, protecting coasts from storm surges. Many divers will have found impressive coral formations and wondered at the colourful fish and crustaceans that inhabit these rich habitats. But few will have ever noticed the subtle variations in colour shown in this photograph of an anemone coral, as the coral polyps emerged at night. In *Another World* (ACC Books, £29.50), photographer Dos Winkel has brought an artist's eye to capture the diversity of colours and textures found on coral reefs.

M.P.



IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Shelf conscious? Many people have no idea that it is the dose that makes a chemical toxic.

Mixed reactions

The Poison Paradox: Chemicals as Friends and Foes

by John Timbrell

Oxford University Press: 2005. 360 pp.
£19.99, \$29.95

Pierluigi Nicotera

It is remarkable that historical periods characterized by a surge of intellectual growth have also been remembered for violence and mischief. As Harry Lime (played by Orson Welles) famously remarked in the film *The Third Man*: "In Italy for 30 years under the Borgias they had warfare, terror, murder and bloodshed, but they produced Michelangelo, Leonardo da Vinci and the Renaissance. In Switzerland they had brotherly love — they had 500 years of democracy and peace, and what did that produce? The cuckoo clock."

Although we may not entirely sympathize with this quote, it is clear that cultural development and scientific and economic progress almost inevitably have side effects. During the Renaissance in France and Italy, the first systematic studies of anatomy and physiology, and the empirical experimentation of the dose–effect relationship (often in humans who did not volunteer), led to a better understanding,

and more accurate use, of poisons.

Unwanted side effects of scientific and technological progress are evident throughout human history. The industrial age has produced one of the most obvious: the contamination of the environment by potentially harmful, synthetic products. More recently, with the mass production of foods, widespread dietary contamination has become a problem. Finally, while a better understanding of disease processes has fostered the development of effective and powerful drugs, these can also cause unexpected, adverse reactions.

Better education and information have also increased the general awareness of the risks posed by chemicals and drugs, raising concern among the public, media and governments. In the past three decades, there has been pressure to reduce environmental contamination, and the safety requirements for foods and drugs have become more stringent. Nevertheless, the public perception of risk from chemicals has since grown disproportionately. It may be possible in many cases to provide safe products without adverse effects, but it is not universally achievable. Hence, the margins for developing safe yet effective medicines have been consistently reduced by the increasing pressure to

develop drugs that are devoid of side effects. For example, aspirin would probably not pass the regulatory process today because of its many side effects, despite its therapeutic value. Chemicals are intrinsically hazardous and all pose some degree of risk, which is perhaps not entirely appreciated by the public. Virtually all human endeavours are potentially harmful, and all agents are potentially toxic. However, the dose or level of exposure and an individual's susceptibility will play a major role in determining the risk posed by a compound.

In *The Poison Paradox*, John Timbrell addresses the problem of the risks posed by chemicals, and considers how, when and why they can be toxic. The book is pervaded by the reminder that it is the dose that makes a chemical toxic. Using easily understandable examples, Timbrell guides the reader through the basic principles of toxicology (such as the interaction between chemicals and biological organisms). Often using an anecdotal style, he clearly explains the dangers and risks linked to natural and synthetic products, including medicines, food and environmental contaminants. Finally, he summarizes the criteria required to assess chemical hazard and risk.

This is not a book for the specialist, but rather is aimed at a general reader who wants to understand the principles that have guided toxicology and the difference between hazards, risks and their assessment. The conclusions are well balanced and illustrate the difficulty of predicting and determining the risk of exposure to agents that we perceive as toxic. One revealing example is given in the chapter on endocrine disruptors — chemicals that cause adverse health effects as a result of changes in hormone function, such as dioxins and oestrogens. Timbrell describes the experimental evidence for their toxicity, the problem with different species' susceptibility to these chemicals, and the lack of convincing evidence for their toxicity in man. He concludes that further evidence is needed to assess their potential adverse health effects in humans.

The chapters on risk assessment and the perception of risk by the public further highlight the need for a better understanding of risk, and show how some risk is unavoidable. Through the analysis of known and yet unresolved problems linked to chemical toxicity, the book also suggests that we need more and diverse research in toxicology. Science, rather than precaution or concern, should drive the assessment of risk and decisions on public health, as Timbrell states: "When the public perception of risk is greater than it really is, politicians need to take steps (which are really unnecessary) to reduce the risk, with the result that huge sums of money may be spent for no real benefit."

The Poison Paradox provides the reader with the information to form an educated perception of risk and its implications in toxicology. ■
Pierluigi Nicotera is at the Toxicology Unit, University of Leicester, Hodgkin Building, Lancaster Road, Leicester LE1 9HN, UK.

An explorer and surveyor

Hermann Weyl made prescient contributions to both mathematics and physics, but also strove to understand reality as a whole.

Frank Wilczek

Hermann Weyl was, according to Fields medallist Michael Atiyah, “one of the greatest mathematicians of the first half of the twentieth century”. Every great mathematician is great in their own way, but Weyl’s way was particularly special. Unlike most modern scientists who choose one or a few specific areas to explore and look neither sideways nor back, Weyl surveyed the whole world. He sought truth and beauty with a discriminating and far-seeing eye.

Weyl’s most unique work is *Philosophy of Mathematics and Natural Science*. No one else could have written it, and no other book I know is like it. I have consulted it many times, and each time I’ve come away enriched. My main purpose here is to direct readers to that book. But since space remains, let me add some background information.

Weyl, who died in 1955 at the age of 70, was a student of David Hilbert at the Georg-August University in Göttingen, Germany, and thus stood in the line of intellectual descent from Carl Gauss, Bernhard Riemann and Lejeune Dirichlet. Upon Hilbert’s retirement, Weyl was invited to take up his chair, but conditions in 1930s Germany and an attractive offer from the new Institute for Advanced Study conspired to bring him to Princeton, where he stayed. With Albert Einstein and John von Neumann, Weyl made the trinity of refugee stars that gave the institute its inimitable scientific lustre.

Einstein and von Neumann had both grown up in the grand German literary and pan-European cultural tradition that was rocked and then shattered by the two world wars. But more than the rebellious Einstein or the protean von Neumann, Weyl embodied that tradition, and his *Philosophy of Mathematics and Natural Science* reflects it in focus, style and erudition.

The main body of text was written in German in 1926, as an article for R. Oldenburg’s *Handbuch der Philosophie*. In 1949, for the English translation, Weyl altered many details and added seven appendices, comprising almost a hundred pages. These centre on relevant scientific events in the intervening years — his passages on Gödel’s theorem, and on quantum mechanics and causality are especially brilliant. But the core of the book had its genesis in the vanished handbook tradition of magisterial



Hermann Weyl touched on questions that are relevant today.

reviews in natural philosophy.

In the introduction, Weyl declares the roots of his literary style: “I was bound by the German literary and philosophical tradition in which I grew up.” As an example of that style, let me quote a passage that I think ranks among the most beautiful and profound passages in all literature: “The objective world simply *is*, it does not *happen*. Only to the gaze of my consciousness, crawling along the lifeline of my body, does a section of this world come to life as a fleeting image in space which continuously changes in time.”

Weyl’s work is also remarkable for its erudition. René Descartes, Gottfried Leibniz, David Hume and Immanuel Kant enter into the discussion as familiar friends. From time to time important literary or philosophical passages are quoted in the original French, German, or even Greek. Weyl’s erudition is not vain display — he is far above that — but a touching assumption of shared culture. Of course the reader will feel at home in this milieu, he seems to assume.

In his biographical memoir of Weyl, Atiyah also talked about Weyl’s work in mathematics, saying: “The last 50 years have seen a remarkable blossoming of just those areas that Weyl initiated. In retrospect, one might almost say that Weyl defined the agenda and provided the proper framework for what followed.” Weyl’s contributions in physics were more sporadic,

but they were significant, and likewise remarkably prescient. In particular, Weyl proposed, and named, the concept of ‘gauge invariance’, which came to dominate fundamental physics during the second half of the twentieth century.

Evidently, Weyl’s intuitions have an excellent track record. By now, parts of the *Philosophy of Mathematics and Natural Science* are dated, of course. Those parts retain interest as intellectual history, because Weyl’s understandings, being close to the best that was possible in their time, serve as benchmarks. But beyond that, *Philosophy of Mathematics and Natural Science*

touches great questions that remain very much alive. Near the end of the book, after a penetrating critique of the concept of causality, Weyl turns to what he calls the body–soul problem — what we know today as the problem of consciousness.

“It is an altogether too mechanical conception of causality that views the mutual effects of body and soul as being so paradoxical that one would rather resort, like Descartes, to the occasionalistic intervention of God or, like Leibniz, to a harmony instituted at the beginning of time.

The real riddle, if I am not mistaken, lies in the double position of the ego: it is not merely an existing individual which carries out real psychic acts, but also ‘vision’, a self-penetrating light (sense-giving consciousness, knowledge, image, or however you may call it); as an individual capable of positing reality, its vision open to reason; a force into which an eye has been put, as Fichte says.”

The fullest enlightenment comes only to those who, like Weyl, discern its absence, seek it, and recognize it when it arrives. ■ Frank Wilczek is at the Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts, 02142, USA.

FURTHER READING

Atiyah, M. in *Biographical Memoirs* Vol. 82, 1–13 (Natl Acad., Washington DC, 2002).
Weyl, H. *Philosophy of Mathematics and Natural Science* (Princeton Univ. Press, Princeton, 1949).

ARCHIVES OF THE INST. OF ADVANCED STUDY

ESSAY

MOLECULAR BIOLOGY

DNA twists and flips

Richard R. Sinden

DNA can shape itself into many forms to achieve its purposes in life. The crystal structure of the junction between two of its forms provides insight into how DNA might accomplish some of these acrobatics.

Crick and Watson's famous structure of DNA has become an icon of our age, somewhat eclipsing DNA's other structural forms. Imagine unravelling the familiar helix (termed B-DNA) and then twisting it up the other way around. The result — Z-DNA — is not so visually appealing, and not so stable, but it has some intriguing biology.

On page 1183 of this issue, Ha *et al.*¹ present the X-ray crystal structure of a junction between B- and Z-DNA. They show that a single base pair is broken and flipped out of the DNA double helix, thereby facilitating the juxtaposition of a left-handed Z-DNA helix with the right-handed B-DNA helix. This deceptively simple observation, however, has ramifications for DNA structure, the dynamics of DNA structural transitions and Z-DNA biology. Nature's design of the B–Z junction maximizes helix integrity and base pairing, thereby minimizing the impact of the dramatic alteration in helix structure where Z-DNA forms in a chromosome.

Z-DNA has seemingly languished for years in the shadows of other DNA structures, despite being one of the first non-B forms to be described². DNA cruciforms, triplexes, quadruplexes and even unwound structures have all been well characterized structurally, and their biological role clearly identified³. Z-DNA can form in certain sequences containing alternating purine and pyrimidine bases, which occur about once every 3,000 base pairs in human chromosomes⁴. Although functions for Z-DNA are still emerging, one role might be to provide a regulatory element to control gene expression. Now the structure of the B–Z junction, 26 years after the discovery of Z-DNA, may revolutionize structural analysis of Z-DNA and help us to elucidate what Z-DNA is doing within the B-DNA that makes up chromosomes.

One reason that the B–Z junction has eluded researchers is that Z-DNA is inherently unstable. To form Z-DNA under physiological conditions, energy from negative DNA supercoiling must be supplied, which is difficult to maintain in a molecule small enough to permit crystallization. In addition, its static appearance in pictures belies the fact that

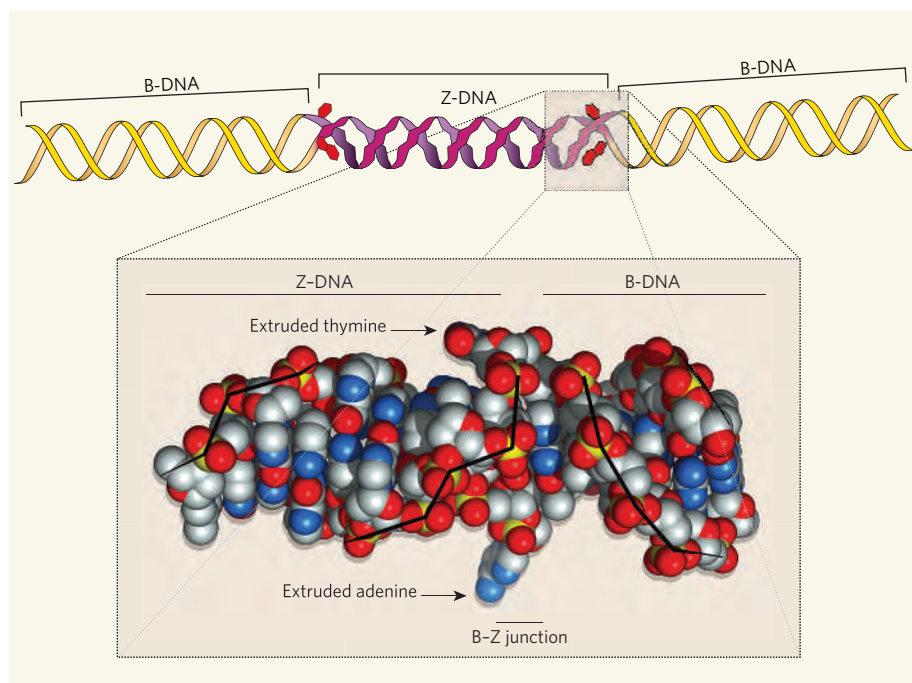


Figure 1 | A new twist. The structure of the B–Z junction solved by Ha *et al.*¹. A region of left-handed Z-DNA is connected to right-handed B-DNA through a junction in which one base pair is flipped out, or extruded, from the DNA helix. (Figure modified from ref. 1.)

DNA is flexible and dynamic⁵. Accordingly, in a supercoiled molecule the two B–Z junctions either side of a Z-DNA tract will probably be mobile and flexible. The discovery of proteins that bind specifically to Z-DNA⁶ implicated this form as a biologically functional DNA component, but it also allowed the crystallographic characterization of the B–Z junction. Ha *et al.*¹ crystallized a 15-base-pair DNA molecule containing a B–Z junction by using a Z-DNA-binding protein. This stabilized an 8-base-pair region of Z-DNA that contained the protein's cognate binding site.

The structure of the B–Z junction comprises a canonical right-handed helix coupled to a typical left-handed Z-DNA helix (Fig. 1). All base pairs within the respective B- and Z-helices retain their standard A–T or C–G base pairing and stacking. In this sense, the B–Z junction is like a four-way cruciform structure in which all bases remain stacked and paired⁷.

The juxtaposition of the left- and right-handed helices, however, is accompanied by the breaking and flipping out of a single A–T base pair at the junction between the two helices (Fig. 1). It is assumed that a C–G base pair might flip out equally well. Extrusion of a base from the helix is a well-known scheme used by enzymes that edit or repair DNA⁸, and which require the base-pairing surface to be exposed for modification. This mechanism has also been observed in an unusual hairpin formed from CCG repeats⁹.

Base extrusion at the B–Z junction is accompanied by a sharp turn in the sugar-phosphate backbone, a slight bend (11°) at the junction, and displacement of the two helical axes by 5.2 Å (ref. 1; Fig. 1). Previously, it was envisaged that there must be a small, unstructured bubble between the opposite-handed helices that could promote unzipping of the helix at the junction. But instead the structure

affords a remarkable conservation of energy and maintenance of helical structure, with only a single base pair broken at the junction.

An important feature of Z-DNA is that, compared with B-DNA, the hydrogen-bonded base pairs are turned upside down. How this structural transition occurs is unclear, but the B–Z junction model may provide a clue. Ha *et al.* suggest that Z-DNA formation might start with the breakage of a single base pair and extrusion of the two bases. These could then swing round 180° and re-pair in the Z-DNA conformation¹. This process would continue from the initiation point, in one or both directions, until an entire tract is transformed into a Z-DNA helix. Once formed, the ends of the Z-DNA tract and the extruded bases may zip back and forth, with dynamic fluctuations in negative superhelical density within a chromosome. This model also suggests a very low energetic barrier to Z-DNA formation, consistent with previous measurements¹⁰.

Z-DNA may provide a sink to absorb the torsional strain left in the wake of a fleeing RNA polymerase or a transient nucleosome, or indeed any other protein complex that interacts dynamically with the DNA. It may also provide a signal for recruitment of an RNA-editing enzyme¹¹ or contribute to expression or repression of genes in different systems^{12–15}. Although additional roles for Z-DNA are expected to emerge, the notion of a Z-DNA helix ensconced within a B-DNA chromosome and punctuated by two pairs of extruded bases presents a more reasonable, stable and energetically feasible view of this key DNA regulatory element. ■

Richard R. Sinden is in the Laboratory of DNA Structure and Mutagenesis, Center for Genome Research, Institute of Biosciences and Technology, Texas A&M University System Health Science Center, 2121 West Holcombe Boulevard, Houston, Texas 77030-3303, USA.
e-mail: rsinden@ibt.tamhsc.edu

CATALYSIS

Gold rush

Masatake Haruta

The chemical industry would be transformed if selective oxidation of hydrocarbons could be achieved efficiently using cheap and clean oxygen from the air. Doing that with gold as a catalyst is a method gaining in allure.

The selective oxidation of hydrocarbons — the targeted addition of oxygen atoms to produce specific desired reaction products — is crucial in industrial petroleum-based chemical processes. Oxygen-containing organic compounds, such as epoxides, ketones, aldehydes, alcohols and acids, are used to produce plastics, detergents, paints, cosmetics and food additives. Oxidation catalysts are second in usage only to polymerization catalysts, accounting for 18% of total catalyst use in the United States in 1991, for example¹. The advancement by Hutchings and colleagues of 'greener' methods for oxidation catalysis using gold (page 1132 of this issue)² is therefore invaluable.

Most industrial oxidation processes involve more than two reaction stages, and tend to use either chlorine or organic peroxides. In the first case, chlorinated organic intermediates are neutralized to form organic oxygenates, producing huge amounts of chloride salts and lesser, but significant, amounts of toxic chlorinated organic by-products. The alternative use of organic peroxides is expensive, and is also accompanied by the formation of by-products — economically disadvantageous if there is no market for them.

So the prospect of using molecular oxygen, O₂, as an oxidizing agent would seem seductive — it is, after all, clean and freely available from air. But the energy required to 'activate' O₂ for reaction by splitting it into its constituent atoms is 498 kilojoules per mol; this is larger than the bonding energy (431 kJ mol^{−1}) of the carbon–hydrogen bond in the most stable hydrocarbon, methane. Controlling the reactivity of atomically dissociated species of oxygen is difficult when such large amounts of energy are put into the system. The most likely course of events during the oxidation of hydrocarbons is that negatively charged oxygen radicals (O[−]) are formed, which associate preferentially with electron-poor carbon atoms, leading to the displacement of hydrogen. In contrast, addition of the oxygen radicals to electron-rich carbon–carbon double bonds (epoxidation) seldom occurs.

All this means that the most common manipulation of hydrocarbons is not selective oxidation but complete oxidation (combustion), producing carbon dioxide and water and generating thermal energy. But a feasible alternative method of controlling the reactivity of oxygen species so as to obtain valuable organic oxygenates is reductive oxygen activation. This technique uses molecular hydrogen or

carbon monoxide to activate oxygen under less extreme conditions, with an energy input of less than 10 kJ mol^{−1}.

Hutchings and colleagues² now show that nanoparticles of gold can help to activate molecular oxygen under mild conditions — at atmospheric pressure and temperatures of 60–80 °C — and so speed up oxidation reactions. Until two breakthrough discoveries in the early 1980s, gold was regarded as having almost no catalytic activity³. But it was then hypothesized, and experimentally confirmed, that the positive gold ion Au³⁺ can catalyse the hydrochlorination of acetylene⁴. Simultaneously, nanoparticles of gold deposited on semiconducting transition-metal oxides were shown⁵ to be surprisingly active in carbon monoxide oxidation, even at −77 °C. Later, gold nanoparticles acting alone were shown to catalyse the selective oxidation of alcohol in water⁶ and the epoxidation of propylene gas using titanium-based oxide supports⁷.

Hutchings and colleagues' use of gold to activate molecular oxygen allows the selective oxidation not just of unsaturated hydrocarbons (generally more reactive owing to the presence of double and triple bonds), but also, in their liquid state, of the less reactive saturated hydrocarbons. In such compounds, all the bonding electrons are used up in single bonds between the constituent atoms. For example, the unsaturated hydrocarbons cyclohexene and *cis*-cyclooctene can be oxidized with 50% and 80% selectivity, respectively. Similarly, the saturated hydrocarbon cyclohexane can be converted into cyclohexanone and cyclohexanol⁸ — both of which form the basis of a wide range of synthetic materials — with a combined selectivity of nearly 100%.

The authors also show that an appropriate choice of solvents allows the catalytic activity of gold to be tuned over a wide range. The product compounds change radically depending on the conditions: whether the reactant has no solvent, or whether water, nonpolar or polar organic solvents are used. All of these solvents seem to require the addition of a small amount of an initiator substance, either hydrogen peroxide or tert-butyl hydroperoxide, to kick-start the reaction. Furthermore, bismuth, which acts as a promoter for palladium and platinum catalysts in the selective oxidation of hydrocarbons, also acts as a promoter for gold. However, the mechanism of oxygen activation probably differs between the two groups of metals.

- Ha, S. C., Lowenhaupt, K., Rich, A., Kim, Y. G. & Kim, K. K. *Nature* **437**, 1183–1186 (2005).
- Wang, A. H. J. *et al.* *Nature* **282**, 680–686 (1979).
- Sinden, R. R. *DNA Structure and Function* (Academic, San Diego, 1994).
- Schroth, G. P., Chou, P.-J. & Ho, P. S. J. *Biol. Chem.* **267**, 11846–11855 (1992).
- Olson, W. K. & Zhurkin, V. B. *Curr. Opin. Struct. Biol.* **10**, 286–297 (2000).
- Rich, A. & Zhang, S. *Nature Rev. Genet.* **4**, 566–572 (2003).
- Lilley, D. M. & Clegg, R. M. *Q. Rev. Biophys.* **26**, 131–175 (1993).
- Roberts, R. J. & Cheng, X. *Annu. Rev. Biochem.* **67**, 181–198 (1998).
- Gao, X., Huang, X., Smith, G. K., Zheng, M. & Liu, H. *J. Am. Chem. Soc.* **117**, 8883–8884 (1995).
- Mirkin, S. M. *et al.* *J. Biomol. Struct. Dyn.* **5**, 79–88 (1987).
- Schwartz, T., Rould, M. A., Lowenhaupt, K., Herbert, A. & Rich, A. *Science* **284**, 1841–1845 (1999).
- Liu, R. *et al.* *Cell* **106**, 309–318 (2001).
- Rothenburg, S., Koch-Nolte, F., Rich, A. & Haag, F. *Proc. Natl Acad. Sci. USA* **98**, 8985–8990 (2001).
- Oh, D. B., Kim, Y. G. & Rich, A. *Proc. Natl Acad. Sci. USA* **99**, 16666–16671 (2002).
- Kim, Y. G. *et al.* *Proc. Natl Acad. Sci. USA* **100**, 6974–6979 (2003).

Gold's catalytic properties seem to depend less on the material on which it rests for liquid-phase oxidation⁹ than they do for gas-phase oxidation¹⁰. However, the support material might be expected to influence at least the dispersion of the gold and thus its catalytic activity. The efficacy of gold nanoparticles as catalysts is also markedly enhanced^{9,11} when the particles are less than 6 nanometres in diameter. Those used by Hutchings and colleagues² were relatively large, with mean diameters of around 25 nanometres, making further improvement in catalytic performance distinctly possible. Future investigations on the effects of size and support material may well reveal a far wider scope for catalysis by gold.

Exactly how gold particles activate molecular oxygen at such low temperatures remains unclear. It is equally uncertain how epoxidation works; this is the oxidation method favoured by industry but which, chemically, is the least probable route to the oxidation of hydrocarbons. Convincing answers to such questions would provide us with a valuable roadmap for pursuing green, sustainable chemistry through control of the reactivity of

oxygen — which is just how enzymes act.

Gold has long been held in the imagination as a thing of never-changing beauty and value. Now it might hold our imagination as an instrument of change at the nanometre scale¹².

Masatake Haruta is in the Materials Chemistry Course, Faculty of Urban Environmental Sciences, Tokyo Metropolitan University, 1-1 Minami-Osawa, Hachioji 192-0397, Tokyo, Japan.

e-mail: haruta-masatake@center.tmu.ac.jp

1. Thayer, A. M. *Chem. Eng. News* **70** (10), 27–49 (1992).
2. Hughes, M. D. *et al.* *Nature* **437**, 1132–1135 (2005).
3. Bond, G. C. & Thompson, D. T. *Catal. Rev. Sci. Eng.* **41**, 319–388 (1999).
4. Hutchings, G. J. *J. Catal.* **96**, 292–295 (1985).
5. Haruta, M., Kobayashi, T., Sano, H. & Yamada, N. *Chem. Lett.* **16**, 405–408 (1987).
6. Prati, L. & Rossi, M. *J. Catal.* **176**, 552–560 (1998).
7. Hayashi, T., Tanaka, K. & Haruta, M. *J. Catal.* **178**, 566–575 (1998).
8. Xu, Y.-J. *et al.* *Catal. Lett.* **101**, 175–179 (2005).
9. Comotti, M., Della Pina, C., Matarrese, R. & Rossi, M. *Angew. Chem. Int. Edn* **43**, 5812–5815 (2004).
10. Haruta, M. *Chem. Rec.* **3**, 75–87 (2003).
11. Tsunoyama, H., Sakurai, H., Negishi, Y. & Tsukuda, T. *J. Am. Chem. Soc.* **127**, 9374–9375 (2005).
12. Hutchings, G. J. & Haruta, M. *Appl. Catal.* **291** (1–2), 1–261 (2005).

EVOLUTION

Along came a sea spider

Graham E. Budd and Maximilian J. Telford

An investigation of brain development in sea spiders provides hints about how the earliest arthropod head evolved. These observations are bound to provoke controversy in an already acrimonious field.

Obscure groups of animals have been making scientific waves lately¹, and few are more obscure than the sea spiders, or pycnogonids. These marine, spider-like animals differ from other arthropods, such as the true spiders, crustaceans and insects, in many ways. Their bodies are so slender that the digestive systems and gonads are squeezed into their limbs; they possess a forward-pointing proboscis with a terminal mouth; and the males brood the eggs. Flanking their unique proboscis is a pair of pincer-bearing appendages known as chelifores, which it has long been assumed are related to the pincer-fangs — chelicerae — of spiders.

Work presented by Maxmen *et al.* on page 1144 of this issue², however, suggests that pycnogonid chelifores and spider chelicerae develop from different regions of the head and therefore cannot be equivalent. At first sight this is a rather esoteric finding. But if it is correct, it will shake up the field of arthropod evolution.

Maxmen *et al.* relied on the fact that each of an arthropod's pairs of appendages is derived from one of the repeating segments that make

up the arthropod body. In addition to its appendages, each segment has a pair of nerve concentrations, or neuromeres. The authors reason that tying the appendages to a specific pair of neuromeres should reveal which segment the appendages belong to. As chelifores and chelicerae are head appendages innervated from the brain, Maxmen *et al.* considered which of the brain neuromeres each appendage is associated with during larval development. Arthropod brains are divided into three regions: protocerebral, deutocerebral and tritocerebral, from front to back. The anterior-most appendage of most living arthropods, including the spider chelicera, is innervated from the deutocerebrum³. What Maxmen *et al.* have now shown, in surprising contrast, is that the pycnogonid chelifores seem to be innervated from the protocerebrum — the most anterior part of the brain. The association of chelifores and chelicerae with different parts of the brain implies that the two types of limb are not equivalent, but are derived from different segments.

This result cuts across previous results based on adult structure⁴, and to see the wider

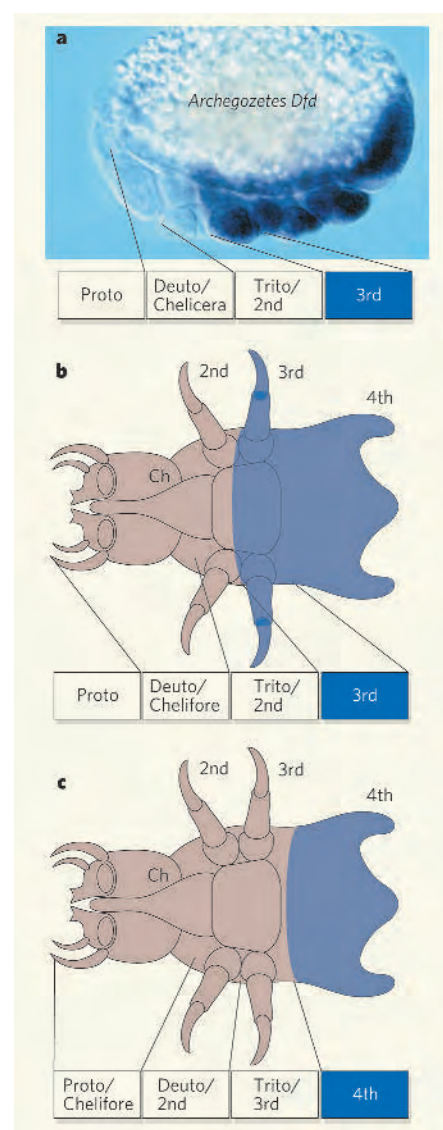


Figure 1 | Testing the results of Maxmen *et al.*² with Hox gene expression. a, Deformed (*Dfd*) expression in an arachnid embryo (the mite *Archegozetes longisetosus*³) compared (b, c) with that expected in a stage III pycnogonid larva under two different reconstructions. b, If the traditionally accepted equivalence of pycnogonid chelifores and arachnid chelicerae is correct, the anterior boundary of pycnogonid *Dfd* will be in the third appendage. c, If Maxmen *et al.* are correct in associating the chelifore with innervation from the protocerebrum, that boundary should be in the fourth appendage. (Appendage innervation: proto, protocerebral; deuto, deutocerebral; trito, tritocerebral. Larval image modified from ref. 10.)

implications we need some historical background. The composition of the arthropod head is one of the bitterest and longest-running problems in animal evolution. Unresolved after more than a century of debate, this sorry tale is (in)famously known as the “endless dispute”⁵.

Much of the attention in this dispute has been directed towards the nervous system. It is widely agreed that the deutocerebrum, tritocerebrum and the more posterior parts of

the nervous system are derived from successive segmental neuromeres. The deutocerebrum innervates the antennae of insects, the anterior antennae of crustaceans and the chelicerae of spiders and scorpions³. But the real meat of the endless dispute has always concerned the nature of the appendage-less front-most part of the brain, the protocerebrum. Is it some sort of non-segmental leftover inherited from the very earliest animal ancestors of the arthropods⁶ (a mystical structure called the acron in the literature), or does it represent the neuromere of a once appendage-bearing anterior segment⁷?

Two lines of evidence have been put forward in support of the existence in ancient arthropods of a protocerebrum with an appendage. First, all extant arthropods (except pycnogonids) possess a small, appendage-like outgrowth of the body which lies just in front of the mouth and is called the labrum. Confusingly, the labrum is not innervated by the protocerebrum (fanning the flames of the dispute); however, in the embryo it starts off right at the front of the animal, and migrates backwards during development. If the labrum represents a highly modified appendage, then its anterior position in development might indicate that it is the long-sought limb of a protocerebral segment.

Second, there is the fossil evidence of the earliest arthropods from 530–490 million years ago. Many of these early arthropods possessed a pair of large, grasping or branched appendages, known as the 'great appendage', found at the anterior of the head. Indeed, a phylogenetic reconstruction published a few years ago suggested that the great appendage was innervated from the protocerebrum⁷. We cannot investigate the nervous system of a fossil, however, and this reconstruction has been hotly disputed, with many researchers preferring to see the great appendage as equivalent to the antennae of insects and crustaceans⁸.

However, if, for the sake of argument, we accept these two lines of evidence at face value, we could reasonably conclude that the protocerebral appendage started out as a great appendage that has subsequently shrunk to the small nub of tissue we now see in most living arthropods as the labrum.

The wider significance of the conclusions of Maxmen *et al.*² now becomes clear. The presence of a bona fide appendage on the pycnogonid protocerebrum (and the absence of a labrum) gives support to the protocerebral origin of the great appendage and to the idea that the labrum is the remnant of this ancient appendage. More excitingly, it implies that the pycnogonids are extraordinary living fossils, retaining an organization of their head that all other living arthropods lost hundreds of millions of years ago.

How, then, might we test these new results²? First, we would like a way to verify the association of chelifore with protocerebrum. One way to achieve this would be to use domains of gene

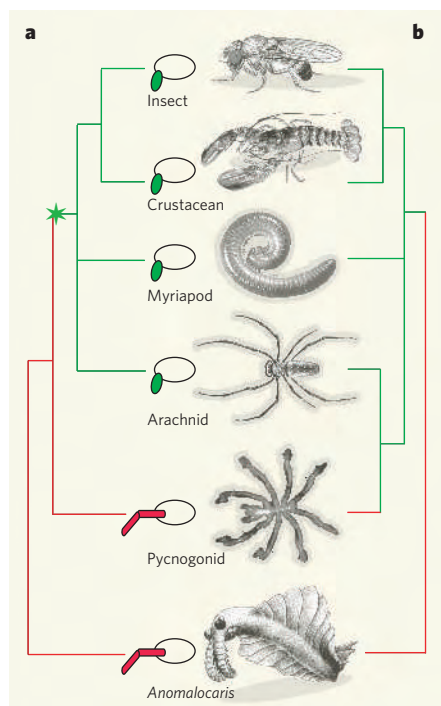


Figure 2 | Possible evolutionary position of pycnogonids and interpretation of Maxmen and colleagues' results². **a**, If pycnogonids branched off before the appearance of insects, crustaceans, myriapods and arachnids, we can interpret their protocerebral chelifores (red) as equivalent to the supposedly anterior great appendage of fossil groups such as *Anomalocaris*. The labrum (green) would have evolved in the common ancestor indicated with a star. **b**, But if pycnogonids are related to arachnids, then either their protocerebral chelifores could be an atavistic re-evolution of the great appendage, or the labrum must have evolved independently in arachnids and the other three taxa. Both of these latter hypotheses are contentious, and could raise doubts about the conclusions of Maxmen and colleagues.

expression as segmental markers. Hox genes are especially useful in this regard, as they have relatively stable domains of expression along the anterior–posterior axis of arthropods. This approach has been used successfully to line up the head segments of arachnids and insects, for example³. Thus, the anteriormost expression of the Hox gene *Deformed* (*Dfd*) marks the fourth segment in both arachnids (Fig. 1a) and insects. Because the arachnid/insect first segment has no associated appendage, this *Dfd* expression in the fourth segment lines up with the third appendage. If the traditional interpretation of pycnogonid appendage assignment is correct, the anteriormost appendage — the chelifore — will be associated with the second segment and, as in arachnids, the fourth-segment expression of *Dfd* will therefore be seen in the third appendage of the pycnogonid larva (Fig. 1b). But if Maxmen *et al.* are correct, the chelifore comes from the first segment and, counting backwards, *Dfd* expression in the fourth segment will therefore be seen in the fourth appendage rather than the third (Fig. 1c).



50 YEARS AGO

It is well known that myxomatosis has caused a high mortality among wild rabbit populations in most parts of England and Wales... Where it has been possible to make accurate observations, a mortality-rate of about 99 per cent was found... Only in one part of the country, the Sherwood Forest area of Nottinghamshire, have attenuated strains of the virus been found in rabbits. In this area estate workers and the Ministry's field-officers noted from April 1955 onwards the presence of an unusual number of individuals that appeared to be recovering... It is emphasized that the attenuation of the myxoma virus in part of Sherwood Forest is an isolated case, affects only a relatively small area, and calls for an increase rather than a slackening of efforts to destroy surviving rabbits.

From *Nature* 22 October 1955.

100 YEARS AGO

Field Book of Wild Birds and their Music — This is a very pretty little book, with many charming illustrations of American singing-birds, and numerous attempts to represent their songs in our musical notation. It would seem as if the songs of American birds lent themselves more readily than those of our European species to such notation, for this is by no means the first attempt of this kind which has recently been made on the other side of the water. The present reviewer is under the disadvantage of not having heard these birds in their native land, and is quite ready to believe that Mr. Mathews's musical notations may give an American the vague idea of what his birds sing; at the same time, as one whose knowledge of music is even older than his knowledge of birds, he must emphatically express a hope that British ornithologists will not imitate their American brethren in trying to render our familiar songs on this system... by far the greater number can only be represented in the amusing way in which Mr. Mathews has noted the song of the bobolink — by a cloudy jumble of notes and lines... which suggests a flute-player gone mad.

From *Nature* 19 October 1905.

50 & 100 YEARS AGO

The second avenue is phylogenetic. The evolutionary scheme we have outlined implies that the transition from great appendage to labrum happened once in the common ancestor of all living arthropods apart from the pycnogonids, which must therefore be very basal in evolutionary terms. But if the pycnogonids truly are the sister group of the spiders and scorpions (which some molecular data suggest⁹), then the results of Maxmen *et al.* will be hard to square (Fig. 2). Testing the phylogenetic position of pycnogonids is therefore crucial.

The conclusions of Maxmen *et al.* overturn entrenched ideas about the body plan of the sea spiders and, furthermore, lend support to some controversial theories of arthropod evolution. Unlike their terrestrial analogues, sea spiders lack a poisonous bite, but this paper is bound to inject venom into what is already one of the most controversial of all zoological topics. ■

Graham E. Budd is in the Department of Earth Sciences, University of Uppsala, Norbyvägen 22, Uppsala SE-752 36, Sweden.

e-mail: graham.budd@pal.uu.se

Maximilian J. Telford is in the Department of Biology, University College London, Gower Street, London WC1E 6BT, UK.

e-mail: m.telford@ucl.ac.uk

1. Bourlat, S. J., Nielsen, C., Lockyer, A. E., Littlewood, D. T. & Telford, M. J. *Nature* **424**, 925–928 (2003).
2. Maxmen, A., Browne, W. E., Martindale, M. Q. & Giribet, G. *Nature* **437**, 1144–1148 (2005).
3. Telford, M. J. & Thomas, R. A. *Proc. Natl Acad. Sci. USA* **95**, 10671–10675 (1998).
4. Winter, G. Z. *Zool. Syst. Evol.* **18**, 27–61 (1980).
5. Rempel, J. G. Q. *Entomol.* **11**, 7–25 (1975).
6. Scholtz, G. *Zoology* **103**, 99–111 (2001).
7. Budd, G. E. *Nature* **417**, 271–275 (2001).
8. Chen, J.-Y., Waloszek, D. & Maas, A. *Lethaia* **37**, 1–17 (2004).
9. Regier, J. C., Shultz, J. W. & Kambic, R. E. *Proc. R. Soc. Lond. B* **272**, 395–401 (2005).
10. Vilpoux, K. & Waloszek, D. *Arthropod Struct. Dev.* **32**, 349–383 (2003).

QUANTUM PHYSICS

Atom waves in passing

Maarten DeKieviet and Joerg Schmiedmayer

Matter-wave interferometers are unique tools for exposing particles acting like waves — one of the stranger facets of quantum theory. They can even measure the quickening of an atom's 'pulse' as it flies past a surface.

Particles sometimes act like waves, and waves sometimes act like particles. This phenomenon, known as wave–particle duality, may seem to confuse what are (to everyday experience at least) two separate and unambiguous concepts. But 100 years after Albert Einstein first introduced the idea of waves behaving like particles to explain the photoelectric effect, and more than 80 years after the French physicist Louis de Broglie proposed the converse behaviour, wave–particle duality has become a staple food of the quantum diet. Writing in *Physical Review Letters*¹, John Perreault and Alexander Cronin expose a further experimental ramification of the effect, by measuring the shift in phase — a wave property — of an atom as it flies past, and interacts with, a surface.

In doing so, they take advantage of matter-wave interferometry, a technique that has in recent decades given fresh impetus to studies of the wave-like nature of particles. Pioneered for simple particles such as electrons² and neutrons (ref. 3 and references therein), the technique has since been extended to larger particles such as atoms and molecules (ref. 4 and references therein).

Interferometry as a generalized technique involves the superposition of two waves to gain information about their relative phase — where in their cycle they are in relation to

each other. A simple analogy is a zip-fastener: for proper zipping, the teeth of one strand must fit perfectly with those of the other. If, however, they are shifted such that the teeth oppose each other, the zipper won't close. Analogously, if the peaks of one wave are next to the troughs of the other, the waves are perfectly out of phase, and their amplitudes cancel out — they interfere 'destructively'. Conversely, if the two waves are perfectly in phase, with the peaks and troughs matching, they interfere constructively to produce a net amplitude that is the sum of the two individual amplitudes.

In an interferometer, an incoming wave is split into two branches. One of these branches is subjected to an outside influence that slows down or speeds up the atom-wave's cycle, or pulse, thus shifting its phase relative to that of the other branch. These two branches are then brought back together and interfere, the amplitude of the resulting wave being proportional to the degree to which the two waves are in phase. Generally in wave mechanics only intensities — the squares of the amplitudes — can be measured, so phase information is lost. The power of interferometry is that it transforms a shift in phase to a change in amplitude, which can be measured as change in intensity.

And so it is in Perreault and Cronin's experiment¹. They make use of a Mach–Zehnder

interferometer, in which a nanoscale grating is used to diffract atomic waves, thus acting as a matter-beam splitter⁵. The authors inserted a further 250-nm-thick membrane with thousands of 50-nm-wide slits into one branch of this interferometer. As they pass through this additional membrane, the atoms experience a weak, attractive van der Waals force through electronic coupling with the membrane's walls. This interaction speeds up the atoms' pulse — the phase of the atom-wave becomes shifted with respect to the free-atom wave in the interferometer's other branch. From the measured interference, the phase shift caused by the atom–surface interaction can be exactly quantified.

This measurable change in the interference pattern arises from an atomic interaction that occurs over a distance up to 1,000 times that of an atomic diameter. Perreault and Cronin are, to their knowledge, the first to determine directly the phase shift caused by the van der Waals interaction between an atom and a surface. The acceleration towards the surface of the channels experienced by the sodium atom-waves is more than a million times that caused by Earth's gravitational field. The channels are, however, very short, so the actual time difference measured by the interferometer is only about 100 attoseconds (10^{-16} seconds). Contrasting this with the overall flight time through the device of around one millisecond gives an idea of the exquisite sensitivity possible with interference experiments. This experiment is a beautiful example of the many tools that are being developed in a true renaissance in the study of atom–surface interactions⁶.

The potential impact of such work stems from its connection to the fields of nanotechnology and atom optics. Nanometre-scale structures could lead to smaller transistors and motors, or the ability to assemble molecules atom by atom. Exploiting the wave behaviour of atoms could lead the way to more precise gyroscopes for navigation, gravity gradiometers for subterranean mapping and other field sensors. The work of Perreault and Cronin¹ lies at the intersection of these two fields, putting a limit on how small nanotechnological and atom-optical devices can be made before the van der Waals interaction disrupts their operation. ■

Maarten DeKieviet and Joerg Schmiedmayer are at the Physikalisches Institut der Universität Heidelberg, Philosophenweg 12, D-69120 Heidelberg, Germany.

e-mail: Schmiedmayer@physi.uni-heidelberg.de

1. Perreault, J. D. & Cronin, A. D. *Phys. Rev. Lett.* **95**, 133201 (2005).
2. Tonomura, A. *et al.* *Electron Holography* (Springer, Berlin, 1999).
3. Rauch, H. & Werner, S. A. *Neutron Interferometry: Lessons in Experimental Quantum Mechanics* (Clarendon, Oxford, 2000).
4. Berman, P. R. (ed.) *Atom Interferometry* (Academic, San Diego, 1997).
5. Keith, D. W. *et al.* *Phys. Rev. Lett.* **66**, 2693–2696 (1991).
6. *Proc. Conf. Atoms and Molecules near Surfaces* Vol. 19; www.iop.org/EJ/toc/1742-6596/19/1 (2005).

CELL BIOLOGY

A BID for the pathway

Michael B. Kastan

Cells have many ways of coping with damage to their DNA, but how are these all coordinated? It seems that BID — a regulator of programmed cell death — stands at the crossroads of several damage-response pathways.

Exposure to DNA-damaging agents can cause mutations, developmental abnormalities or cancer, and cells have developed numerous ways to minimize these effects. Such mechanisms include cell-cycle checkpoints to prohibit damaged cells from dividing while the cell deals with the damage; processes to repair the DNA; and programmed cell death (apoptosis). Although it makes sense that there should be coordinated regulation of these different processes, they have mostly been studied as independent pathways. A new link between these mechanisms comes from work reported in *Cell* by Zinkel *et al.*¹ and Kamer *et al.*² demonstrating an unexpected role for the protein BID, a known regulator of apoptosis, in the control of cell-cycle progression following DNA damage.

Apoptosis can be initiated by external signals from other cells — TNF- α , for instance, which is released by certain immune cells to kill infected cells — or by internal warnings resulting from DNA damage or other cellular stresses. Regardless of the stimulus, the death process is controlled by one of several signalling pathways, all of which include one or more members of the BCL-2 family of proteins. The binding of external factors to their cognate 'death receptors' on the cell surface leads to the cleavage and activation of BID³ — a BCL-2 family member. BID, in turn, activates BAX and BAK, both of which are BCL-2 proteins that then trigger apoptosis (Fig. 1)⁴. By contrast, DNA damage causes activation of the ATM or ATR protein kinases (enzymes that phosphorylate certain target proteins), and this activation leads to a rise in p53 protein levels⁵. This protein can switch on a number of target genes, including those encoding the BCL-2 family members BAX, PUMA and NOXA, each of which contributes to apoptosis in specific cell types⁶. In certain cell types or physiological settings, however, the induction of p53 protein, instead of initiating apoptosis, causes arrest of the cell cycle at the G1 phase. Moreover, other targets of ATM and ATR also contribute to cell-cycle arrest in the S and G2/M phases⁵.

Previously, the p53-dependent activation of the genes encoding BAX, PUMA and NOXA

was thought to be the main link between DNA damage and the control of apoptosis. Now, Zinkel *et al.*¹ and Kamer *et al.*² show that another pathway links the control of apoptosis to that of the cell-cycle arrest following DNA damage. Both groups found that BID is phosphorylated in an ATM-dependent manner after DNA damage and that some of the BID protein then moves into the nucleus. Unexpectedly, both groups also found that arrest of

The importance of BID and its phosphorylation by ATM in influencing cell survival following DNA damage varied in different experiments in the two papers. These discrepancies may have resulted from differences in cell types, agents or doses used, but they need to be explained if we are to understand the physiological roles of this branch of the ATM pathway. Furthermore, there were inconsistencies between the two studies regarding which DNA damaging agents activate the ATM-BID pathway. For example, Zinkel *et al.*¹ found that ATM was required for BID phosphorylation following exposure to many different types of agents, including hydroxyurea and ultraviolet light, whereas Kamer *et al.*² observed BID phosphorylation and ATM-dependence only after treatment with agents that introduce breaks in DNA strands, such as ionizing irradiation and the drug etoposide. The specificity of ATM-dependence for responses to ionizing irradiation would be consistent with a large amount of data in the literature⁵.

Although there are some confusing results, the two papers convincingly establish that the apoptotic protein BID is a target of ATM and that it has an unanticipated role in controlling cell-cycle progression following DNA damage. The data demonstrate that BID and ATM come together in one arm of one pathway, but also underscore the notion that the two proteins have other distinctive biochemical and functional roles. Thus, it is not surprising that the effects of their loss are quite different: mice lacking BID develop a blood disorder resembling chronic myelomonocytic leukaemia⁷, whereas mice and humans lacking ATM develop acute lymphoid leukaemias and lymphomas⁸. It makes sense that cells would have a multi-pronged, coordinated response to stresses such as DNA damage, because it would be advantageous to simultaneously control DNA repair, cell-cycle progression and programmed cell death.

BID is unlikely to be the last example of a protein that stands at a crossroads to influence multiple parts of the stress response.

Michael B. Kastan is in the Department of Hematology-Oncology, St Jude Children's Research Hospital, Room D-5048, 332 North Lauderdale Street, Memphis, Tennessee 38105, USA.
e-mail: Michael.Kastan@stjude.org

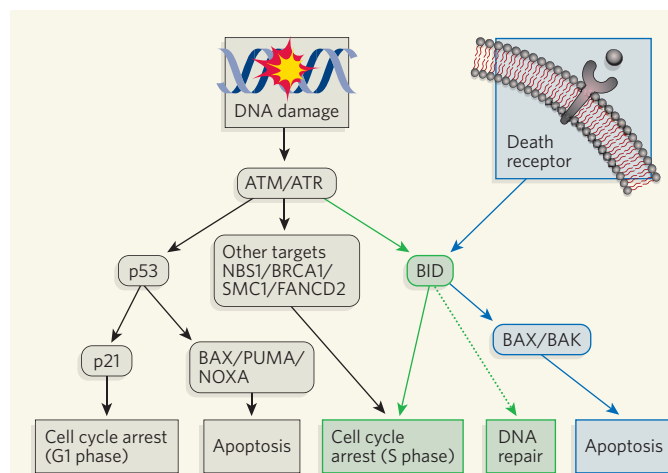


Figure 1 | BID linking two pathways. Zinkel *et al.*¹ and Kamer *et al.*² find an unexpected role for BID in the ATM pathway that responds to DNA damage (new steps in green). Following DNA damage, ATM and ATR are activated and, through p53, can cause either cell-cycle arrest at the G1 stage or apoptosis. ATM and ATR also activate several other protein targets, causing the cell cycle to stall in the S phase. Now BID can be added to this list of ATM/ATR targets. Although BID clearly induces apoptosis following activation of death receptors, its role in controlling apoptosis following DNA damage needs to be clarified. The dotted arrow shows a link that is suggested from Zinkel and colleagues' results, but has yet to be demonstrated.

the cell cycle in S phase in response to DNA damage requires BID phosphorylation. Moreover, the domains of the BID protein responsible for control of the cell cycle are distinct from that involved in its apoptotic function. These observations add BID to a surprisingly long list of proteins that are phosphorylated by ATM or ATR to control the progression through the cell cycle after DNA damage (Fig. 1). Furthermore, Zinkel *et al.*¹ demonstrate that cells lacking BID suffer more chromosomal aberrations after DNA damage than those that have BID, raising the possibility that this protein participates in repair of the DNA as well as in cell-cycle control and apoptosis.

Although the two papers agree in several aspects, there are some apparent discrepancies and some findings that require further clarifi-

1. Zinkel, S. S. *et al. Cell* **122**, 579–591 (2005).
2. Kamer, I. *et al. Cell* **122**, 593–603 (2005).
3. Zha, J. *et al. Science* **290**, 1761–1765 (2000).
4. Wang, X. *Genes Dev.* **15**, 2922–2933 (2001).
5. Kastan, M. B. & Bartek, J. *Nature* **432**, 316–323 (2004).
6. Danial, N. N. & Korsmeyer, S. J. *Cell* **116**, 205–219 (2004).
7. Zinkel, S. S. *et al. Genes Dev.* **17**, 229–239 (2003).
8. Shiloh, Y. *Nature Rev. Cancer* **3**, 155–168 (2003).

ASTRONOMY

Odd company

Fulvio Melia

Black holes cannot yet be seen directly, but their influence on surrounding stars is allowing them to be identified with increasing certainty. That those stars are there to be influenced, though, raises other questions.

Observations of stellar dynamics have so far revealed compact, dark masses at the centre of almost 40 galaxies¹. These are commonly assumed to be supermassive black holes — collapsed massive objects whose gravitational pull is so great that no light or matter can escape them. In most cases, however, what is seen is also consistent with the dark masses being simply clusters of dark stars. The exceptions are the dark masses at the core of the galaxy NGC 4258, in our own Milky Way and now, as Bender *et al.*² report in *The Astrophysical Journal*, in our near neighbour the Andromeda galaxy.

Without actually seeing the dark pit it creates by absorbing or bending all the light incident upon it^{3,4}, the most compelling method to prove the existence of a black hole is to constrain its size and mass. If a very massive object is confined to a compact region within a critical 'Schwarzschild' radius dictated by the general theory of relativity, its gravitational pull so warps space-time that this wraps round to enclose the body, preventing anything escaping. To fulfil this criterion and so be considered a black hole, the object of three million solar masses that lurks at the centre of our Galaxy, for example, must be five times smaller than Mercury's orbit around the Sun.

To measure whether a dark mass of unknown provenance is a black hole, one must first find an object moving under its influence: in newtonian mechanics, the speed of an orbiting object, together with its radius from the central source of gravity, is sufficient to determine that source's mass. The extreme gravitational pull of a black hole makes it difficult to find objects near it, let alone measure their speed. But if a close-orbiting object can be found, it can be used to rule out other potential identities for the central mass concentration: if the orbital radius is smaller than the size of other distributions of matter, such as a neutrino ball or a cluster of dark, dead stars, the only remaining viable possibility for the gravitational source is a black hole.

The nucleus of Andromeda comprises a central dark-matter distribution and — as we now know, thanks to remarkable observations from the Hubble Space Telescope — not one, but three concentrations of starlight. Two of these, commonly labelled P1 and P2 (peaks 1 and 2), were known previously⁵. Conventional wisdom⁶ has it that they are merely the opposite ends of an elliptical distribution of old stars orbiting the central distribution, which is near P2 at one focus of the ellipse. Bender and

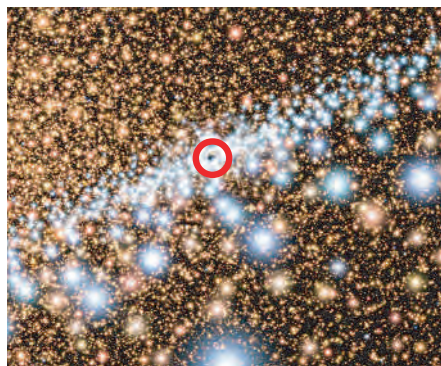


Figure 1 | Black and blue. An artist's impression of the mysterious young, blue stars in the region known as P3 encircling the supermassive black hole at the nucleus of the Andromeda galaxy. The region of older, redder stars, called P2, that surrounds P3 is typical of the cores of galaxies. The black hole itself (circled in red) appears as a distorted shadow at the centre of the blue disk. The 400 young, blue stars apparently formed in a burst of activity about 200 million years ago — posing problems for existing theories of star formation around supermassive objects.

colleagues², however, now confirm the existence of a third stellar component, P3, a tiny nucleus of hot, blue stars embedded within P2 (Fig. 1). An unusual blue concentration in P2 had been noted earlier⁷, but the fact that it constitutes a compact disk of stars, separate from P2, has been established only with the latest observations. Ironically, stars such as these have no business being so close to a black hole — yet, following the reasoning above, their existence there rules out any other explanation for the concentration of mass in Andromeda's nucleus other than it's being a black hole.

Despite its small size (barely a light year across), P3 contains stars with the highest average circular rotation velocity — almost 1,700 kilometres per second — measured so far in any galaxy. According to Newton's law of gravity, the central mass required to corral such fast stars so close to the nucleus exceeds that of 100 million Suns, rendering Andromeda's black hole at least 30 times bigger than its counterpart at the heart of the Milky Way. For other dark objects, such as brown dwarfs (stars that have failed to ignite) or dead stars, to mimic such a single massive object, more than 100 million of them would have to be concentrated within a region only a third of a light year across. The collisions that would ensue would destroy this structure in only a few million years, so it would

not have lasted long enough for us to see it.

But there remains the problem of the age of the roughly 400 stars that form P3. They are bright and blue, and therefore very young compared with the age of the Andromeda galaxy; most of them must have formed less than 200 million years ago. So did these stars form *in situ*, or did they migrate from farther out? Given their short lifespan, it is very difficult to see how they could have diffused inwards (and still be visible as young stars now) through two-body interactions. The alternative is that they formed where they are now, through the collapse of infalling molecular clouds. But the gravity of the black hole, and the extreme physical conditions at the centre of Andromeda, would greatly inhibit star formation unless the cloud density were many orders of magnitude greater — thus facilitating gravitational condensation into stars — than is generally encountered there.

The situation of the stars orbiting the black hole at the centre of our own Galaxy is also bizarre^{8,9}. Young stars, less than 10 million years old, are assembled there into two counter-rotating disks; closer still to the centre, some 20 stars orbit within only a few light days of the black hole, some at speeds that can exceed 5,000 kilometres per second. One possible explanation for these close, young stars is that two clouds of matter fell into the black hole together, each colliding and compressing the gas of the other such that the material could clump together, thus overcoming the many factors that would otherwise inhibit their contraction into stars⁹. Although improbable, this may explain what we see in the Galactic Centre; it is less likely to account for the single disk of stars seen in P3 at the nucleus of Andromeda.

Thus, although the unknown dark-mass concentrations at the nucleus of many galaxies are conforming to what is becoming the 'standard model' of black holes¹⁰, other mysteries of similar opacity are emerging. As Bender *et al.* note, there is no plausible explanation of how and why the hot, young stars near the centre of the Milky Way and Andromeda got there. It seems that only a new theory of star formation in the chaotic environment surrounding a supermassive object will suffice. For the positive identification of Andromeda's centre, black-hole enthusiasts are thankful nonetheless. ■

Fulvio Melia is in the Departments of Physics and Astronomy, The University of Arizona, Tucson, Arizona 85721, USA.

e-mail: melia@physics.arizona.edu

1. Kormendy, J. in *Coevolution of Black Holes and Galaxies* (ed. Ho, L. C.) 1–20 (Cambridge Univ. Press, 2004).
2. Bender, R. *et al. Astrophys. J.* **631**, 280–300 (2005).
3. Falcke, H. *et al. Astrophys. J.* **528**, L13–L17 (2000).
4. Bromley, B. C. *et al. Astrophys. J.* **555**, L83–L87 (2001).
5. Lauer, T. R. *et al. Astron. J.* **106**, 1436–1437 (1993).
6. Tremaine, S. *Astron. J.* **110**, 628–633 (1995).
7. King, I. R. *et al. Astron. J.* **109**, 164–172 (1995).
8. Ghez, A. *et al. Nature* **407**, 349–351 (2000).
9. Genzel, R. *et al. Astrophys. J.* **594**, 812–832 (2003).
10. Melia, F. *The Edge of Infinity: Supermassive Black Holes in the Universe* (Cambridge Univ. Press, 2003).

NASA, ESA & A. SCHALLER (FOR STSC)

EVOLUTIONARY BIOLOGY

Fruitfly genome is not junk

Alexey S. Kondrashov

A comparison of two fruitfly genomes shows that much of their non-coding DNA is controlled by either negative or positive selection, dealing a double blow to the neutral theory of molecular evolution.

Once upon a time, the world seemed simple when viewed through the eyes of evolutionary biologists. All genomes were tightly controlled by various forms of natural selection. DNA encoded functional genes, and most mutations that occurred were rejected through negative selection. Those exceptional mutations that were beneficial substituted for the original gene variant (allele) and spread through the evolving populations by positive selection. And polymorphisms — where several alleles coexist within a population — were maintained by yet another, balancing, form of selection.

This idyllic world begun to crumble in 1968, when Kimura¹ made his modest proposal that most allele substitutions and polymorphisms do not substantially affect an organism's fitness and are governed, not by positive or balancing selection, but by random drift. Kimura still allowed for negative selection to eliminate most new mutations, so this proposal can be regarded as 'weak neutralism'. However, a decade later the onset of large-scale genome sequencing led to the discovery of genes that were so degraded as to be no longer functional (pseudogenes) and of other junk DNA. This led to 'strong neutralism', which claims that regions of genomes that do not encode proteins consist mostly of functionless DNA, ignored by all forms of selection². Indeed, current estimates of the fraction of functionally important segments in mammalian non-coding sequences range from 10–15% (ref. 3) to just 3% (ref. 4).

On page 1149 of this issue, however, Andolfatto³ reports a strikingly pre-neutralist pattern for a sample of fruitfly genomes, those of *Drosophila melanogaster* and *Drosophila simulans*. First, he compared non-coding nucleotide sites with synonymous sites — protein-coding sites where, because of the redundancies in the nucleotide triplet code, a substitution would not alter the encoded amino acid. The synonymous sites were used as a paradigm of neutrality, for want of a better one. Andolfatto found reduced levels of interspecies divergence and of intraspecies polymorphism within *D. melanogaster*, suggesting that around 50% of non-coding sites in *Drosophila* are affected by negative selection more strongly than synonymous sites. So because synonymous sites are also subject to some negative selection in *Drosophila*⁶, most of the fly's non-coding sequences must be under some functional constraint.

Second, a substantial fraction of those

nucleotide substitutions that do occur in non-coding *Drosophila* sequences are driven by positive selection. This conclusion follows from the results of the McDonald–Kreitman test, a statistical analysis that detects positive selection acting on certain kinds of nucleotide sites from the excess of substitutions, relative to polymorphisms, at these sites⁷. So, if Andolfatto's results are confirmed by further genome-scale analyses, neither strong nor even weak neutralism describes the evolution of the *D. melanogaster* and *D. simulans* genomes.

Can the neutral theory survive this double blow? Easily, because, unlike the fruitfly genomes, mammalian genomes are certainly full of junk. Although some originally junk sequences can be recruited to perform a function, and so become subject to selection^{8,9}, there is little doubt that substitutions and polymorphisms are, indeed, effectively neutral in the bulk of mammalian non-coding DNA.

At least two classes should therefore be recognized among the genomes of multicellular eukaryotes, which have long non-coding regions. Negative selection in mammals, and more generally in vertebrates, is so weak or inefficient (presumably because of their low effective population sizes) that even long segments of junk DNA often spread through the population. As a result, bloated and mostly neutral mammal-like genomes evolve.

By contrast, genomes of *D. melanogaster* and its close relatives (although not all of *Drosophila*), and probably those of many other species, are protected from the rampant accumulation of junk DNA by efficient selection. In *D. melanogaster*, individual transposable elements (jumping DNA segments) are mostly kept at low frequencies¹⁰, introns (segments of non-coding DNA within genes) are comparatively short, and pseudogenes are few¹¹. In such species, selection maintains lean and mostly functional melanogaster-like genomes.

Although the neutral theory will survive Andolfatto's demonstration that there is pervasive selection in *Drosophila*, its original justification may not. Kimura¹ claimed that most substitutions must be neutral because positive selection driving all of them would incur too high a fitness cost. However, it is now known that the 'lag load'¹² associated with even rapid adaptive evolution is not necessarily very high¹³. So Andolfatto's conclusion that one selection-driven substitution has occurred about every ten generations since

D. melanogaster and *D. simulans* diverged seems reasonable theoretically. If flies can do it, so could others, invalidating Kimura's original argument.

In *Drosophila*, the relatively junk-free regions between genes, which probably regulate gene expression, seem to be a major target of positive selection. It is therefore fair to assume that the accumulation of beneficial mutations also had a major role in the evolution of functional segments in the intergenic regions in mammal-like genomes. However, nature is queerer than we may think. Since the human–chimpanzee divergence, functionally important intergenic segments have incorporated many deleterious mutations, rather than beneficial ones, perhaps because of the relatively recent decline of the effective population size in hominids¹⁴. We do not yet know how quickly beneficial mutations accumulate in functional segments of mammal-like genomes in the lineages where efficiency of selection did not decline.

The total number of functionally important nucleotides in the genome, T , is crucially important for estimating the genomic deleterious mutation rate U , a key parameter in evolutionary genetics¹⁵. Indeed, in a diploid organism, $U = 2T\mu$, and μ , the per nucleotide mutation rate, can be measured directly. Andolfatto's data and analysis suggest that, when U is estimated for *D. melanogaster*, μ must be multiplied by about 2×10^8 . Recently, $\mu = 2 \times 10^{-8}$ was reported in the nematode worm *Caenorhabditis elegans*¹⁶, and analogous data for *Drosophila* are expected soon. These data will have profound implications for a variety of outstanding problems in evolutionary biology, in particular with regard to the evolution of sex. It is truly amazing how little we know quantitatively about mutation and selection in the genomes of even the most well-studied organisms. ■

Alexey S. Kondrashov is at the National Center for Biotechnology Information, National Institutes of Health, Bethesda, Maryland 20894, USA.

e-mail: kondrashov@ncbi.nlm.nih.gov

- Kimura, M. *Nature* **217**, 624–626 (1968).
- Li, W.-H., Gojobori, T. & Nei, M. *Nature* **292**, 237–239 (1981).
- Shabalina, S. A., Ogurtsov, A. Y., Kondrashov, V. A. & Kondrashov, A. S. *Trends Genet.* **17**, 373–376 (2001).
- Mouse Genome Sequencing Consortium. *Nature* **420**, 520–559 (2002).
- Andolfatto, P. *Nature* **437**, 1149–1152 (2005).
- Akashi, H. *Genetics* **144**, 1297–1307 (1996).
- Smith, N. G. & Eyre-Walker, A. *Nature* **415**, 1022–1024 (2002).
- Jordan, I. K. et al. *Trends Genet.* **19**, 68–72 (2003).
- Silva, J. C. et al. *Genet. Res.* **82**, 1–18 (2003).
- Bartolomé, C., Maside, X. & Charlesworth, B. *Mol. Biol. Evol.* **19**, 926–937 (2002).
- Harrison, P. M. et al. *Nucl. Acids Res.* **31**, 1033–1037 (2003).
- Maynard Smith, J. *The Evolution of Sex* (Cambridge Univ. Press, 1978).
- Kondrashov, A. S. *J. Theor. Biol.* **107**, 249–260 (1984).
- Keightley, P. D., Lercher, M. J. & Eyre-Walker, A. *PLoS Biol.* **3**, e42 (2005).
- Charlesworth, B. & Charlesworth, D. *Genetica* **102/103**, 3–19 (1998).
- Denver, D. R. et al. *Nature* **430**, 679–682 (2004).

BRIEF COMMUNICATIONS

Intercepting the first rat ashore

Keeping islands free of rats may be harder with isolated invaders as these evade conventional trapping.

A single Norway rat released on to a rat-free island was not caught for more than four months, despite intensive efforts to trap it. The rat first explored the 9.5-hectare island and then swam 400 metres across open water to another rat-free island, evading capture for 18 weeks until an aggressive combination of detection and trapping methods were deployed simultaneously. The exceptional difficulty of this capture indicates that methods normally used to eradicate rats in dense populations are unlikely to be effective on small numbers, a finding that could have global implications for conservation on protected islands.

Invasive rodent species severely disrupt island ecosystems^{1,2}. Although rats can be eradicated from islands^{2,3}, they often reinvade: at least 11 New Zealand islands have been reinvaded since 1980 by Norway rats (*Rattus norvegicus*)⁴. It is difficult to eliminate rats in the early stages of invasion^{3,5}. This may be related to unusual behaviour by the animals, because the problems exceed those expected from the proportionally lower probability of detection due to low population density⁶. This behaviour could be monitored by releasing individual, radio-collared rats on to rat-free islands⁷.

The uninhabited and forested Noises Islands (Motuoropapa and Otata) off northeast New Zealand have been reinvaded by Norway rats at least six times between 1981 and 2002 (ref. 4); however, by 2004 they had been rat-free for more than two years. In November 2004, we captured an adult male Norway rat by using a chocolate-baited trap (Table 1, method 1) on the uninhabited and forested Pakihi Island, which lies 30 km to the southeast. The rat was



Space invader: rats can swim between islands.

fitted with a radio collar and a DNA sample taken from its tail before it was released on a beach on Motuoropapa (area, 9.5 hectares). Our aim was to study the behaviour of this solitary invading rat and to test its susceptibility to standard methods of detection (independently of radio-tagging) and elimination. (For details of methods, see supplementary information.)

The rat traversed the entire island before settling after 4 weeks in a home range of about a hectare, as defined by radiotelemetry. All attempts from weeks 4 to 8 to capture or detect the rat using conventional techniques failed (Table 1, methods 2–5). This failure was surprising because devices 2, 3 and 4 had been laid at double the density that had been used to eliminate previous populations⁸.

After 10 weeks, the radio signal was lost. Fresh signs of the rat subsequently appeared in a 100-metre grid of bait stations and tracking tunnels (Table 1, method 6) on Otata (area, 21.8 hectares), which lies 400 m across open water⁸. Genetic fingerprinting of rat faeces found on this island confirmed that this was the same individual as the one released on Motuoropapa (results not shown).

From weeks 12 to 16, different techniques were used to try to eliminate the rat on the second island (Table 1, methods 5–9). Positive signs using methods 5–7 showed that it had visited various sites. However, the rat was not

caught until 18 weeks after its release, when it was finally killed in a trap that had been baited with fresh penguin and set where trained dogs had found concentrated rodent scent.

Norway rats can supposedly swim up to 600 m (ref. 1) but, to our knowledge, this is the first record of a rat swimming hundreds of metres across open water. Although Motuoropapa had previously supported 4.2 rats per hectare and contains forest plants, invertebrates and sea birds as uncontested food sources⁸, the rat still swam to another island, possibly prompted by the absence of conspecifics.

Field studies of rat dispersal have previously focused on high-density populations, from which juveniles escape because of competition for food and territory^{9,10}; however, new invaders may behave differently⁵. Our findings confirm that eliminating a single invading rat is disproportionately difficult, not only because of atypical behaviour in the absence of conspecifics, but also because bait can be less effective in the absence of competition for natural food sources. Our results may help in the design of conservation strategies to keep islands free of invasive rodents.

James C. Russell*†, David R. Towns‡, Sandra H. Anderson*, Mick N. Clout*

*School of Biological Sciences and †Department of Statistics, University of Auckland, Private Bag 92019, Auckland, New Zealand
e-mail: j.russell@auckland.ac.nz

‡Department of Conservation, Private Bag 68 908, Newton, Auckland, New Zealand

1. Atkinson, I. A. E. in *Conservation of Island Birds* (ed. Moors, P. J.) 35–81 (International Council for Bird Preservation, Technical Publication No. 3, Cambridge, 1985).
2. Simberloff, D. *Trends Ecol. Evol.* **16**, 273–274 (2001).
3. Towns, D. R. & Broome, K. G. N. *Z. J. Zool.* **30**, 377–398 (2003).
4. Clout, M. N. & Russell, J. C. in *Proc. Int. Conf. on Assessment and Control of Biological Invasion Risks* (eds Iwatsuki, K. & Urano, K.) (IUCN, Gland, Switzerland and Cambridge, UK, in the press).
5. Thorsen, M., Shorten, R., Lucking, R. & Lucking, V. *Biol. Cons.* **96**, 133–138 (2000).
6. Holway, D. A. & Suarez, A. V. *Trends Ecol. Evol.* **14**, 328–330 (1999).
7. Dilks, P. J. & Towns, D. R. *DOC Sci. Int. Ser.* **59** (Dept Conservation, Wellington, New Zealand, 2002).
8. Moors, P. J. N. *Z. J. Ecol.* **8**, 37–54 (1985).
9. Lacey, E. A. & Solomon, N. G. *J. Mammal.* **84**, 1135–1140 (2003).
10. Macdonald, D. W., Mathews, F. & Berdoy, M. in *Ecologically Based Rodent Management* (eds Singleton, G. R., Hinds, L. A., Leirs, H. & Zhang, Z.) 49–80 (ACIAR, Canberra, 1999).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.
doi:10.1038/4371107a

Table 1 | Attempts to detect and eliminate a rat

Method	Number	Exposure (h)
1 Live trap	30	10
2 Snap trap	20	70
3 Waxed device	15	70
4 Tracking tunnel	15	70
5 Trained dogs	2	16
6 Permanent grid	20	250
7 Peanut butter bait	20	50
8 Buried traps	5	40
9 Poison bait	5	30

Different strategies were used intensively on two islands to track and trap a single rat after its initial capture (by method 1) and release: methods 2–5 were used on Motuoropapa; methods 5–9 were used on Otata. Further details of the devices and detection methods are available in supplementary information.

ANALGAMATED PHOTOS: T. FUKAMIE & G. Y. SASAKI

AVIANFLU

Isolation of drug-resistant H5N1 virus

The persistence of H5N1 avian influenza viruses in many Asian countries and their ability to cause fatal infections in humans have raised serious concerns about a global flu pandemic¹. Here we report the isolation of an H5N1 virus from a Vietnamese girl that is resistant to the drug oseltamivir², which is an inhibitor of the viral enzyme neuraminidase and is currently used for protection against and treatment of influenza. Further investigation is necessary to determine the prevalence of oseltamivir-resistant H5N1 viruses among patients treated with this drug.

An H5N1 influenza virus, A/Hanoi/30408/2005, was isolated on 27 February 2005 from a 14-year-old Vietnamese girl (patient 1) who had received a prophylactic dose (75 mg once a day) of oseltamivir from 24 to 27 February and was given a therapeutic dose (75 mg twice daily) for 7 days starting on 28 February. No virus was isolated from specimens after the administration of increased doses of oseltamivir. The patient recovered and was discharged from hospital on 14 March 2005.

Direct sequencing after amplification by polymerase chain reaction of the virus isolated from a specimen collected on 27 February indicated that some of the virus population had a histidine-to-tyrosine substitution at position 274 (represented as H274Y) in its neuraminidase protein, a mutation that confers resistance to oseltamivir^{3–5}. We therefore tested the sensitivity of the virus to oseltamivir carboxylate⁶ (the active form of the drug) and

found that the dose required for 50% inhibition of neuraminidase activity (IC_{50}) in the isolate was 90 nM, which exceeds the IC_{50} for oseltamivir-sensitive viruses (0.1–10 nM)⁷. We then plaque-purified the virus.

Ten viral clones, randomly picked from the resultant plaques, were classified into three groups according to their response to oseltamivir (see supplementary information): six were highly resistant to the drug ($IC_{50} > 763$ nM), three were slightly resistant (IC_{50} values between 7.1 and 12.5 nM) and one was highly sensitive ($IC_{50} = 0.6$ nM). The highly resistant viruses had tyrosine at position 274 in their neuraminidase, whereas those showing only slight resistance had serine at position 294.

Patient 1 had not had any known direct contact with poultry, but had cared for her 21-year-old brother (patient 2) while he had a documented H5N1 virus infection (for details of the disease course and treatment in these patients, see supplementary information). We found that the neuraminidase gene of the brother's virus was identical to clone 7 of the girl's virus (see supplementary information). Also, the haemagglutinin gene of the brother's virus was identical to clones 2 and 9 of the girl's virus, apart from a nucleotide change at position 271. The timing of infection in these two patients, together with the lack of known interaction of the girl with poultry, raises the possibility that the virus could have been transmitted from brother to sister.

We assessed the growth of a highly oseltamivir-resistant clone (H274Y, clone 9) and of an oseltamivir-sensitive clone (H274, clone 7) in ferrets⁸. Viral titres were higher in animals infected with the oseltamivir-sensitive virus (Fig. 1a; $P = 0.000099$; two-way repeated measures ANOVA). Oseltamivir treatment reduced viral titres in animals infected with the drug-sensitive virus (Fig. 1b; $P = 0.048$, Student's *t*-test), but not in animals infected with the resistant virus (Fig. 1b; $P = 0.23$, Student's *t*-test). However, all of the viral clones, including those highly resistant to oseltamivir, were sensitive to zanamivir¹⁰ (IC_{50} , 0.5–3.1 nM), another neuraminidase inhibitor. In ferrets, we found that zanamivir treatment reduced viral titres in animals infected with virus that was oseltamivir-sensitive (Fig. 1b; $P = 0.0000019$, Student's *t*-test) or oseltamivir-resistant (Fig. 1b; $P = 0.018$, Student's *t*-test).

We investigated how the viruses

bound *in vitro* to different configurations of sialyl glycopolymers, similar to those on the host's cell-surface receptor¹¹. We compared binding by two of the viral clones (clones 7 and 9) with binding by an avian flu virus (A/duck/Mongolia/301/2001) and another human flu virus (A/Kawasaki/1/2001). We found that both H5N1 clones bound to α -2,3-linked polymer and (less efficiently) to α -2,6-linked polymer (Fig. 1c, clone 7; clone 9, not shown). The A/duck/Mongolia/301/2001 virus also bound α -2,3-linked polymer but did not bind α -2,6-linked polymer at all; the A/Kawasaki/1/2001 virus bound strongly to α -2,6-linked polymer but only weakly to α -2,3-linked polymer (results not shown). The broader binding properties of our H5N1 viral clones may reflect a degree of adaptation in human hosts.

Although our findings are based on a virus from only a single patient, they raise the possibility that it might be useful to stockpile zanamivir as well as oseltamivir in the event of an H5N1 influenza pandemic. They also highlight the importance of monitoring the emergence of drug resistance in H5N1 isolates from patients treated with neuraminidase inhibitors.

Q. Mai Le*, Maki Kiso†‡, Kazuhiko Someya†§, Yuko T. Sakai†‡, T. Hien Nguyen*, Khan H. L. Nguyen*, N. Dinh Pham*, Ha H. Nguyen||, Shinya Yamada†‡, Yukiko Muramoto†‡, Taisuke Horimoto†‡, Ayato Takada†‡, Hideo Goto†‡, Takashi Suzuki†‡¶, Yasuo Suzuki†‡¶, Yoshihiro Kawaoka†‡¶#☆

*National Institute of Hygiene and Epidemiology, Hanoi, Vietnam

†Division of Virology, Department of Microbiology and Immunology, #International Research Center for Infectious Diseases, Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan

e-mail: kawaoka@ims.u-tokyo.ac.jp

‡Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, Kawaguchi, Saitama 332-0012, Japan

§Daichi Pharmaceutical, Tokyo 134-8630, Japan

||National Institute for Clinical Research in Tropical Medicine, Hanoi, Vietnam

¶Department of Biochemistry, University of Shizuoka, School of Pharmaceutical Sciences and COE Program in the 21st Century, Yada, Shizuoka-shi 422-8526, Japan

☆Department of Pathobiological Sciences, School of Veterinary Medicine, University of Wisconsin, Madison, Wisconsin 53706, USA

1. Enserink, M. *Science* **309**, 370–371 (2005).
2. Treanor, J. J. et al. *J. Am. Med. Assoc.* **283**, 1016–1024 (2000).
3. Gubareva, L. V., Kaiser, L. & Hayden, F. G. *Lancet* **355**, 827–835 (2000).
4. Gubareva, L. V., Kaiser, L., Matrosovich, M. N., Soo-Hoo, Y. & Hayden, F. G. *J. Infect. Dis.* **183**, 523–531 (2001).
5. Zambon, M. & Hayden, F. G. *Antiviral Res.* **47**, 1–17 (2000).
6. Potier, M., Mameli, L., Belisle, M., Dallaire, L. & Melancon, S. B. *Anal. Biochem.* **94**, 287–296 (1979).
7. Hurt, A. C., Barr, I. G., Gunter, H. & Hampson, A. W. *Antiviral Res.* **62**, 37–45 (2004).
8. Toms, G. L. R., Bird, R. A., Kingsman, S. M., Sweet, C. & Smith, H. *Br. J. Exp. Pathol.* **57**, 37–48 (1977).
9. Varghese, J. N., Laver, W. G. & Colman, P. M. *Nature* **303**, 35–50 (1983).
10. Varghese, J. N., McKimm-Breschkin, J. L., Caldwell, J. B., Kortt, A. A. & Colman, P. M. *Proteins* **11**, 49–56 (1992).
11. Shinya, K. et al. *J. Virol.* **79**, 9926–9932 (2005).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

doi:10.1038/4371108a

Published online 14 October 2005.

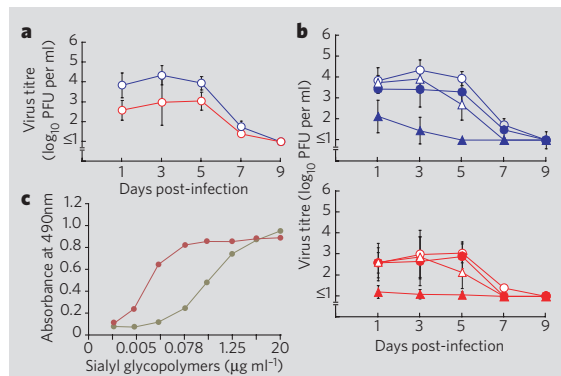


Figure 1 | Drug and binding sensitivity of H5N1 viral clones isolated from a human patient. **a, b**, Viral titres from ferrets (5 animals per group) infected with H5N1 viruses. PFU, plaque-forming units. **a**, Comparison of titres from oseltamivir-sensitive (blue) and oseltamivir-resistant (red) viruses in mock-treated animals (error bars, s.d.). **b**, Comparison of titres in animals treated with oseltamivir or zanamivir (error bars, s.d.). Top, oseltamivir-sensitive virus; bottom, oseltamivir-resistant virus: open symbols, mock-treated; filled circles, oseltamivir-treated; filled triangles, zanamivir-treated. **c**, Receptor specificity of viral clone 7 from patient 1, showing binding activity to sialyl glycopolymers containing either α -2,3-linked (red) or α -2,6-linked (brown) sialic acids. The maximum standard deviation of the data points shown was less than 0.09. For details of methods, see supplementary information.

The CREB coactivator TORC2 is a key regulator of fasting glucose metabolism

Seung-Hoi Koo^{1*}, Lawrence Flechner^{1*}, Ling Qi¹, Xinmin Zhang¹, Robert A. Screaton¹, Shawn Jeffries¹, Susan Hedrick¹, Wu Xu², Fayçal Boussouar², Paul Brindle², Hiroshi Takemori³ & Marc Montminy¹

Glucose homeostasis is regulated systemically by hormones such as insulin and glucagon, and at the cellular level by energy status. Glucagon enhances glucose output from the liver during fasting by stimulating the transcription of gluconeogenic genes via the cyclic AMP-inducible factor CREB (CRE binding protein). When cellular ATP levels are low, however, the energy-sensing kinase AMPK inhibits hepatic gluconeogenesis through an unknown mechanism. Here we show that hormonal and energy-sensing pathways converge on the coactivator TORC2 (transducer of regulated CREB activity 2) to modulate glucose output. Sequestered in the cytoplasm under feeding conditions, TORC2 is dephosphorylated and transported to the nucleus where it enhances CREB-dependent transcription in response to fasting stimuli. Conversely, signals that activate AMPK attenuate the gluconeogenic programme by promoting TORC2 phosphorylation and blocking its nuclear accumulation. Individuals with type 2 diabetes often exhibit fasting hyperglycaemia due to elevated gluconeogenesis; compounds that enhance TORC2 phosphorylation may offer therapeutic benefits in this setting.

Type 2 diabetes affects nearly 20 million individuals in the United States alone^{1,2}. Diabetes complications such as renal failure, micro-vascular disease and peripheral neuropathy account for an increasing proportion of annual health care costs. Tight glycaemic control has been associated with a reduced incidence of diabetes complications, underscoring efforts to characterize regulatory pathways that function importantly in glucose metabolism³.

Circulating glucose levels reflect a balance between glucose production by the liver and glucose utilization by skeletal muscle. Under fasting conditions, this balance is largely maintained by decreasing glucose uptake into muscle and increasing glucose output from the liver via gluconeogenesis. Pancreatic glucagon normally triggers the activation of catabolic programmes in the liver in part via the cAMP responsive factor CREB^{4–6}. CREB in turn stimulates hepatic gluconeogenesis as well as fatty acid oxidation by inducing expression of the nuclear hormone receptor coactivator PGC-1 α (peroxisome-proliferation-activated receptor- γ coactivator-1)^{7–10}. Mice deficient in PGC-1 α , due either to acute RNA interference (RNAi)-mediated knockdown or to targeted gene disruption of the *PGC-1 α* gene, display fasting hypoglycaemia with reduced expression of gluconeogenic genes^{9–11}.

Under feeding conditions, insulin silences hepatic PGC-1 α expression and transcriptional activity via the Akt-mediated phosphorylation and nuclear export of the forkhead family activator FOXO1 (refs 6, 12). Indeed, loss of insulin signalling in diabetes leads to elevated hepatic PGC-1 α expression and excessive hepatic glucose production⁸. Within this regulatory framework the gluconeogenic programme is further modified by adipocyte-derived hormones called adipokines^{13,14}. Among these, resistin and adiponectin have been shown to modulate gluconeogenesis via the energy-sensing kinase AMPK, although the underlying mechanisms have not been determined.

Role of CREB coactivators during fasting

Glucagon is thought to enhance hepatic gene expression by stimulating the protein kinase A (PKA)-mediated phosphorylation of CREB at Ser 133 and by promoting the subsequent recruitment of the coactivator CBP (CREB-binding protein) to the promoter^{15,16}. To test this notion, we injected mice intraperitoneally with glucagon or insulin, and monitored CREB activation in liver. Intraperitoneal injection of glucagon stimulated hepatic CREB phosphorylation (p-CREB) within 10 min, as detected by histochemical and western blot analysis (Fig. 1a, b). Unexpectedly, intraperitoneal insulin administration had comparable effects on p-CREB levels (Fig. 1a, b). To rule out potential counter-regulatory effects of hypoglycaemia after insulin treatment, we co-injected glucose (0.5 g per kg body weight) with insulin and observed comparable CREB phosphorylation relative to insulin alone (Supplementary Fig. S1). Taken together, these results indicate that the CREB–CBP pathway is unlikely to discriminate between fasting and feeding signals in liver.

We next tested the potential involvement of other CREB coactivators in this setting. In recent studies, cAMP has been found to stimulate cellular gene expression by triggering the dephosphorylation and nuclear entry of TORCs, a family of latent cytoplasmic coactivators that enhances cellular gene expression via an interaction with the CREB basic region/leucine zipper (bZIP) DNA-binding domain^{17–20}. We examined TORC2 activity because this family member was expressed at highest levels relative to TORC1 and TORC3 in liver (data not shown).

Adenovirus-expressed TORC2 (Ad-TORC2) was localized primarily in the cytoplasm under *ad libitum* feeding conditions, translocating to the nucleus within 10 min after intraperitoneal glucagon administration (Fig. 1a). By contrast with its effect on CREB phosphorylation, insulin did not promote Ad-TORC2 nuclear entry, demonstrating the capacity of TORC2 to distinguish between

¹Peptide Biology Laboratories, Salk Institute for Biological Studies, 10010 N Torrey Pines Rd, La Jolla, California 92037-1002, USA. ²Department of Biochemistry, St Jude Children's Research Hospital, 332 N. Lauderdale, Memphis, Tennessee 38105, USA. ³Department of Biochemistry and Molecular Biology, Graduate School of Medicine (H-1), Osaka University, 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan.

*These authors contributed equally to this work.

fasting and feeding cues. Consistent with this activation profile, both Ad-TORC2 and endogenous TORC2 were highly phosphorylated at Ser 171 under *ad libitum* feeding or insulin-stimulated conditions, and were dephosphorylated only in response to glucagon or fasting (Fig. 1b; see also Supplementary Figs S1 and S2). Arguing against an inhibitory role for insulin itself in modulating TORC2 localization

directly, endogenous hepatic TORC2 was comparably dephosphorylated in mice injected with glucagon or glucagon plus insulin (Supplementary Fig. S1).

After nuclear entry, TORC2 is thought to stimulate cellular gene expression via an association with CREB on relevant promoters¹⁹. In chromatin immunoprecipitation assays of liver tissue samples, glucagon but not insulin promoted recruitment of TORC2 to gluconeogenic genes (Fig. 1c). Similarly, addition of the cAMP agonist forskolin (FSK) to primary cultures of rat hepatocytes induced TORC2 occupancy over *PEPCK* (phospho-enol pyruvate carboxykinase), *G6Pase* (glucose-6-phosphatase) and *PGC-1 α* promoters (Fig. 1d). FSK also stimulated CREB phosphorylation over gluconeogenic genes, indicating that both CREB–TORC and CREB–CBP pathways are probably activated in this setting (Supplementary Fig. S3). Consistent with this notion, we observed CBP recruitment to *PEPCK* and *G6Pase* promoters in response to FSK (Supplementary Fig. S3).

Although CREB activity is regulated by both CBP and P300, CBP has been proposed to exert a key role in modulating hepatic gluconeogenesis; hepatic glucose output during feeding can be blocked by insulin-dependent phosphorylation of CBP at a site (Ser 436) that is absent in P300 (ref. 21). To test the importance of

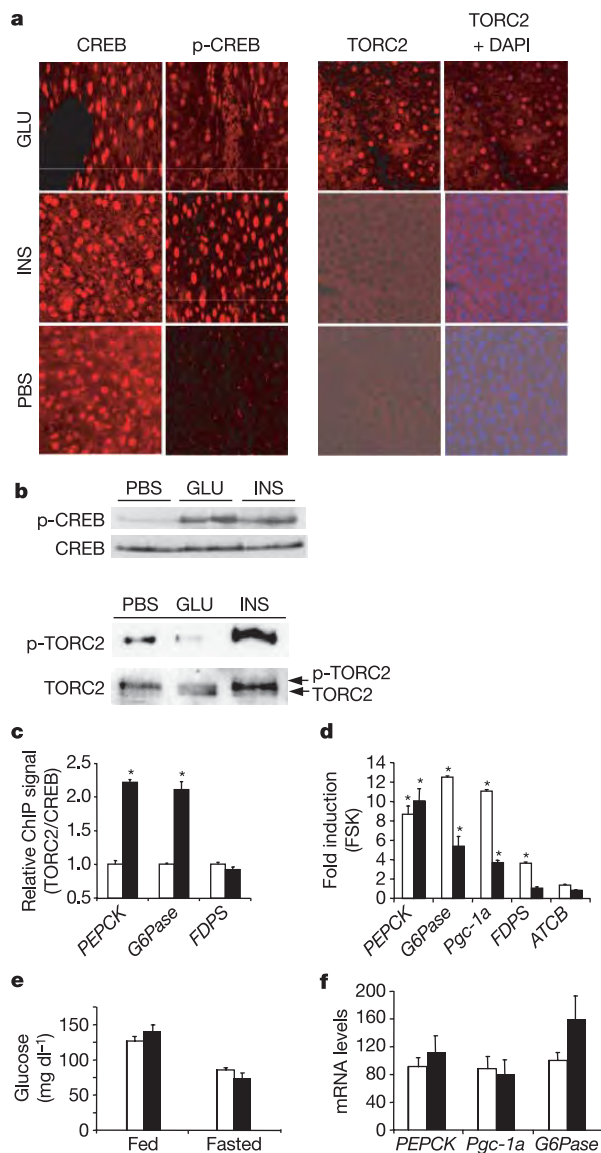


Figure 1 | Fasting hormones activate TORC2. **a**, **b**, Immunohistochemical (**a**) and western blot (**b**) assays showing effect of insulin (INS), glucagon (GLU) or saline (PBS) injection (10 min, intraperitoneal injection) on CREB, phospho-CREB (p-CREB), TORC2 and phospho-TORC2 (p-TORC2) localization and levels in liver. **c**, Relative effect of insulin (open columns) and glucagon (filled columns) on TORC2 occupancy over *PEPCK* and *G6Pase* genes relative to CREB by chromatin immunoprecipitation (ChIP) assay (TORC2/CREB; asterisk, $P < 0.05$, t -test). **d**, Effect of FSK on p-CREB (open columns) and TORC2 (filled columns) occupancy over gluconeogenic and control (*FDPS* and *ATCB*) promoters in primary hepatocytes (asterisk, $P < 0.05$, t -test). **e**, **f**, CREB–CBP complexes are not required for glucose homeostasis. **e**, Blood glucose levels for *CBP*^{+/+} (open columns; $n = 10$) and *CBP*^{kix/kix} (filled columns; $n = 7$) mice under fed ($P = 0.19$) and fasted (18 h, $P = 0.19$) conditions (t -test). **f**, Quantitative PCR analysis of *PEPCK* ($P = 0.43$), *Pgc-1 α* ($P = 0.83$) and *G6Pase* ($P = 0.11$) mRNA levels in livers from fasted *CBP*^{+/+} (open columns; $n = 7$) and *CBP*^{kix/kix} (filled columns; $n = 7$) mice. mRNA levels in this and other figures are shown as relative values. Data in **c–f** represent mean $\pm$ s.e.m.

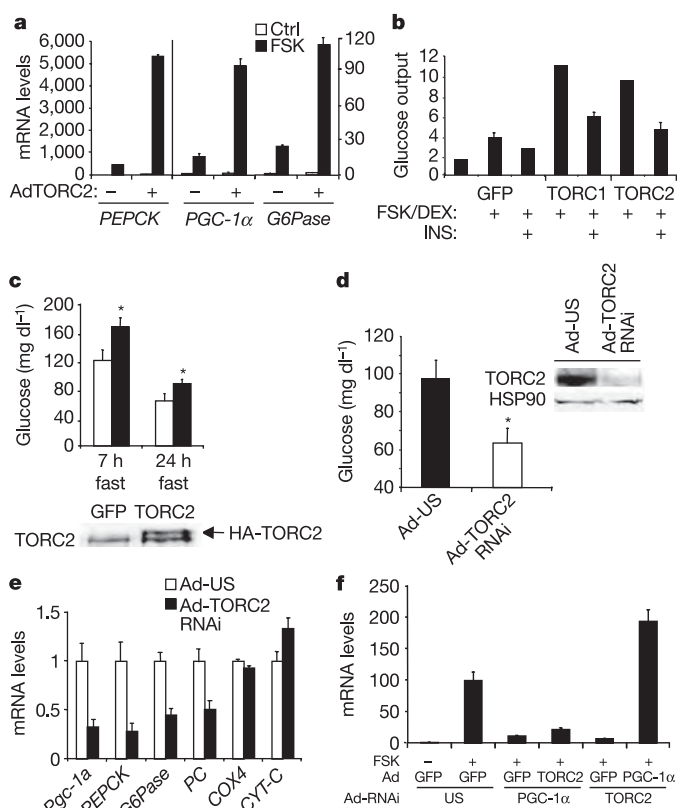


Figure 2 | TORC2 is required for fasting hepatic gluconeogenesis.

a, **b**, Effect of Ad-TORC2 on gluconeogenic gene expression (**a**) and glucose output (**b**) in primary hepatocytes exposed to FSK, FSK plus dexamethasone (FSK/DEX) and insulin, as indicated. **c**, Blood glucose levels in Ad-GFP (open columns) and Ad-TORC2 (filled columns) mice after 7-h and 24-h fasts (asterisk, $P < 0.01$, t -test) ($n = 3$). Bottom: western blot showing levels of Ad-TORC2 (HA-TORC2) relative to endogenous TORC2 in livers from mice above. **d**, **e**, Effect of Ad-TORC2 RNAi or control (Ad-US) RNAi on fasting glucose levels ($n = 3$) (asterisk, $P < 0.01$, t -test) (**d**) and gluconeogenic gene expression (**e**). Inset in **d** shows a western blot of hepatic TORC2 levels in Ad-US and Ad-TORC2 RNAi mice. **f**, Effect of Ad-PGC-1 α and Ad-TORC2 on *PEPCK* mRNA levels in primary hepatocytes co-expressing Ad-TORC2 RNAi, Ad-PGC-1 α RNAi, or Ad-US. Data in **a–f** represent mean $\pm$ s.e.m.

the CREB–CBP interaction in this process we used *CBP^{kix/kix}* mice (*CBP* is also known as *Crebbp*), which contain knockin mutations in the CREB-binding domain (KIX; amino acids 586–672) that specifically block CREB–CBP complex formation²². If insulin disrupts gluconeogenesis by blocking the CREB–CBP interaction, then *CBP^{kix/kix}* mice would be expected to display fasting hypoglycaemia and reduced gluconeogenic gene expression. Remarkably, blood glucose levels were very similar between *CBP^{kix/kix}* mice and control littermates under either feeding or fasting conditions (Fig. 1e). Moreover, we observed no discernable difference in fasting levels of *PEPCK* (also known as *Pck2*) *G6Pase* (also known as *G6pc*) or *Pgc-1 α* (also known as *Ppargc1a*, the product of which is PGC-1 α) messenger RNAs between *CBP^{kix/kix}* and wild-type mice (Fig. 1f), arguing against the notion that recruitment of CBP to CREB target genes *per se* is necessary for induction of the gluconeogenic programme.

Role of TORC2 in glucose metabolism

On the basis of the ability of TORC2 to enter the nucleus and to associate with CREB target genes in response to fasting signals, we explored the role of this coactivator in directing the gluconeogenic programme. Ad-TORC2 had marginal effects on gluconeogenic genes under basal conditions, but strongly potentiated their expression after exposure to FSK (Fig. 2a). The stimulatory effects of TORC2 were CREB-dependent; expression of a dominant negative CREB polypeptide (A-CREB), which specifically inhibits binding of CREB but not other bZIP family members to DNA²³, disrupted TORC2 activity on *PEPCK*, *G6Pase* and *PGC-1 α* genes (Supplementary Fig. S4 and data not shown).

The ability of TORC2 to stimulate the gluconeogenic programme

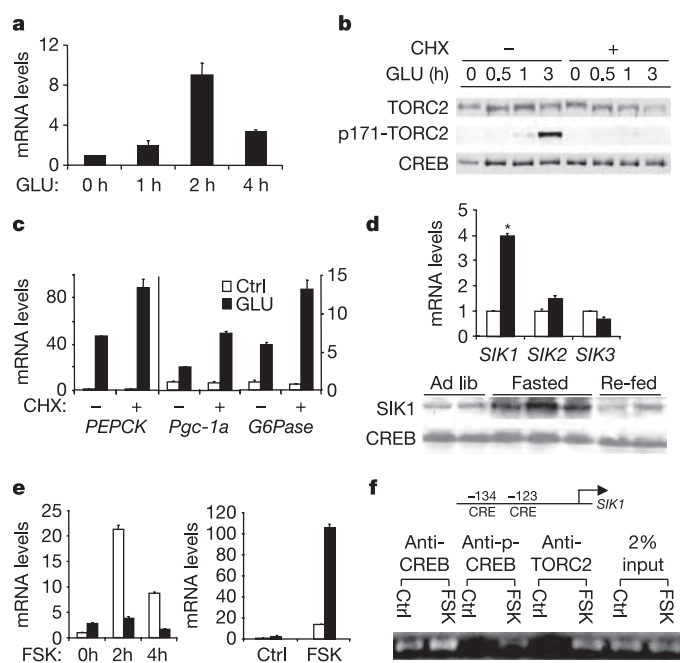


Figure 3 | Induction of SIK1 during fasting attenuates the gluconeogenic programme in primary hepatocytes. **a**, Time course of *PEPCK* gene expression after exposure to glucagon. **b**, **c**, Effect of glucagon and CHX on TORC2 or p-TORC2 protein levels (**b**), and on mRNA levels for gluconeogenic genes (**c**). **d**, Effect of feeding, fasting, or re-feeding on *SIK1*, *SIK2* and *SIK3* mRNA levels (top, $n = 3$) (asterisk, $P < 0.01$, t -test) and on *SIK1* protein levels in liver (bottom). Control, open columns; fasted, filled columns. **e**, Effect of Ad-GFP (open columns) or Ad-A-CREB (filled columns) (left) and Ad-GFP (open columns) or Ad-TORC2 (filled columns) (right) on *SIK1* mRNA levels in primary hepatocytes exposed to FSK. **f**, Top: CREs on *SIK1* promoter. Bottom: chromatin immunoprecipitation assay of CREB, p-CREB and TORC2 on *SIK1* promoter from control or FSK-treated (10 μ M, 2 h) cells. Data in **a**, **c**–**e** represent means $\pm$ s.e.m.

in hepatocytes suggests that it may promote glucose output in response to fasting signals. Ad-TORC2 enhanced glucose production in primary rat hepatocytes after exposure to FSK plus dexamethasone; these effects were inhibited by co-treatment with insulin (Fig. 2b; see also Supplementary Fig. S5). When expressed at levels comparable to endogenous TORC2 in liver, Ad-TORC2 also promoted fasting hyperglycaemia *in vivo* (Fig. 2c). Indeed, levels of circulating insulin were elevated in Ad-TORC2 mice, indicating that the enhanced hepatic glucose output in this setting is sufficient to trigger a counter-regulatory response (Supplementary Fig. S6).

On the basis of its ability to stimulate the gluconeogenic programme and to promote hyperglycaemia when overexpressed in mice, endogenous TORC2 may be required for the metabolic response to fasting signals in liver. To test this idea, we used an Ad-TORC2 RNAi construct to reduce hepatic TORC2 expression (80% by western blot assay; Fig. 2d). Mice made acutely deficient in TORC2 exhibited fasting hypoglycaemia (65 mg dl⁻¹ versus 100 mg dl⁻¹); hepatic mRNA levels for gluconeogenic genes were reduced 2–4-fold relative to control littermates (Fig. 2e). Taken together, these results indicate that TORC2 exerts significant regulatory control over hepatic gluconeogenesis in the fasted state.

In previous studies, the nuclear hormone receptor coactivator PGC-1 α was found to direct expression of the gluconeogenic programme in liver during fasting^{7–10}. Although the ability of TORC2 to stimulate PGC-1 α expression and to associate with its promoter suggests that TORC2 lies upstream of PGC-1 α , our results did not exclude a scenario in which TORC2 and PGC-1 α individually regulate fasting metabolism in liver via parallel pathways. In that case, overexpression of one coactivator (TORC2 or PGC-1 α) might be expected to compensate for deficiencies in the other. Knockdown of either PGC-1 α or TORC2 with Ad-RNAi constructs disrupted *PEPCK* gene expression in primary hepatocytes exposed to FSK (Fig. 2f). Ad-PGC-1 α rescued expression of the *PEPCK* gene in TORC2-deficient cells, but Ad-TORC2 had no effect on *PEPCK* expression in PGC-1 α -deficient cells (Fig. 2f), demonstrating that PGC-1 α probably acts downstream of TORC2 in the gluconeogenic pathway.

Attenuation of TORC2 by a feedback loop

Similar to other second messengers, cAMP stimulates cellular gene expression with burst-attenuation kinetics²⁴. After exposure of primary hepatocytes to glucagon, *PEPCK* mRNA levels became maximal after 2 h and returned to near baseline after 4 h (Fig. 3a). Consistent with this profile, TORC2 was re-phosphorylated at Ser 171, and hence inactivated, 3 h after glucagon stimulation in primary rat hepatocytes (Fig. 3b). Pre-treatment with the protein synthesis inhibitor cycloheximide (CHX) blocked the phosphorylation of TORC2 at Ser 171 by glucagon, suggesting that fasting signals promote the synthesis of an inhibitor, which in turn feeds back to shut down the CREB–TORC pathway (Fig. 3b). Supporting this notion, CHX pre-treatment enhanced gluconeogenic gene expression 2–3-fold in primary rat hepatocytes exposed to glucagon (Fig. 3c).

On the basis of the ability of CHX to disrupt TORC2 Ser 171 phosphorylation during the attenuation period, we reasoned that glucagon may induce the expression of an inhibitory kinase. In previous studies, the SIK (salt-inducible kinase; also known as SNF1-like kinase) family of AMP kinases was found to associate with and to phosphorylate TORC2 at Ser 171 (ref. 19), an optimal consensus site for phosphorylation by AMPK family members (LNRTSSDSAL; Ser 171 is underlined). Indeed, fasting increased *SIK1* mRNA and protein levels in liver fourfold relative to feeding conditions, leaving other SIK family members (*SIK2* and *SIK3*) relatively unaffected (Fig. 3d). Similarly, exposure of primary rat hepatocytes to glucagon or FSK stimulated *SIK1* mRNA and protein accumulation via a CREB-dependent mechanism; Ad-TORC2 potentiated *SIK1* gene expression in cells exposed to FSK (Fig. 3e; see also Supplementary Fig. S7). Notably, the accumulation of SIK1 protein in cells exposed

to glucagon was blocked by CHX, implicating SIK1 as an inducible silencer of gluconeogenic genes (Supplementary Fig. S7).

In studies to determine how fasting signals regulate SIK1 expression, we noticed two consensus cAMP-responsive promoter elements (CREs) in rat, mouse and human orthologues of the SIK1 promoter (Fig. 3f). In transient assays of human HepG2 hepatocytes, PKA stimulated *SIK1* promoter activity about 20-fold; these effects were disrupted by co-expression of A-CREB (Supplementary Fig. S8). Indeed, CREB was found to occupy the *SIK1* promoter in chromatin immunoprecipitation assays of primary rat hepatocytes; TORC2 was recruited to this promoter in response to FSK treatment, indicating that fasting signals regulate SIK1 gene expression via the CREB–TORC pathway (Fig. 3f).

SIK1 inhibits hepatic gluconeogenesis

If SIK1 functions as part of a negative feedback loop for the gluconeogenic programme, then reducing cellular SIK1 levels would be expected to enhance the expression of these genes by increasing TORC2 activity. Levels of unphosphorylated—and there-

fore activated—TORC2 were twofold higher in Ad-SIK1 RNAi-expressing hepatocytes compared to control cells after exposure to glucagon (Fig. 4a; see also Supplementary Fig. S9); we observed a corresponding increase in mRNA levels for *PEPCK*, *Pgc-1 α* and *G6Pase* in SIK1-deficient cells, confirming the inhibitory role of this kinase on the gluconeogenic programme (Fig. 4b). Indeed, knockdown of SIK1 in mice promoted both fasting hyperglycaemia and gluconeogenic gene expression, demonstrating a key role for this kinase *in vivo* (Fig. 4c, d).

On the basis of the ability of Ad-SIK1 RNAi to enhance gluconeogenesis, we suspected that, conversely, SIK1 overexpression would block induction of the gluconeogenic programme. When expressed in hepatocytes, SIK1 promoted Ser 171 phosphorylation and hence disruption of TORC2 activity on the *PEPCK* promoter in cells exposed to FSK (Supplementary Figs S10 and S11, left panel). By contrast, SIK1 had no effect on the ability of PGC-1 α to stimulate transcription from a PPAR- α target gene (acyl CoA oxidase (*AOX*)), arguing against an inhibitory effect of this kinase on gluconeogenic regulators downstream of TORC2 (Supplementary Fig. S11, right panel).

Having seen that SIK1 inhibits TORC2 activity on gluconeogenic promoters, we evaluated its role in fasting glucose metabolism. Relative to control Ad-GFP littermates, Ad-SIK1 as well as Ad-SIK2 mice exhibited fasting hypoglycaemia and reduced gluconeogenic gene expression (Fig. 4e, f). Although the ability of Ad-SIK1 to block hepatic glucose output in this setting could reflect an unanticipated induction of the insulin signalling pathway, fasting insulin levels were actually lower in Ad-SIK1 and Ad-SIK2 compared to

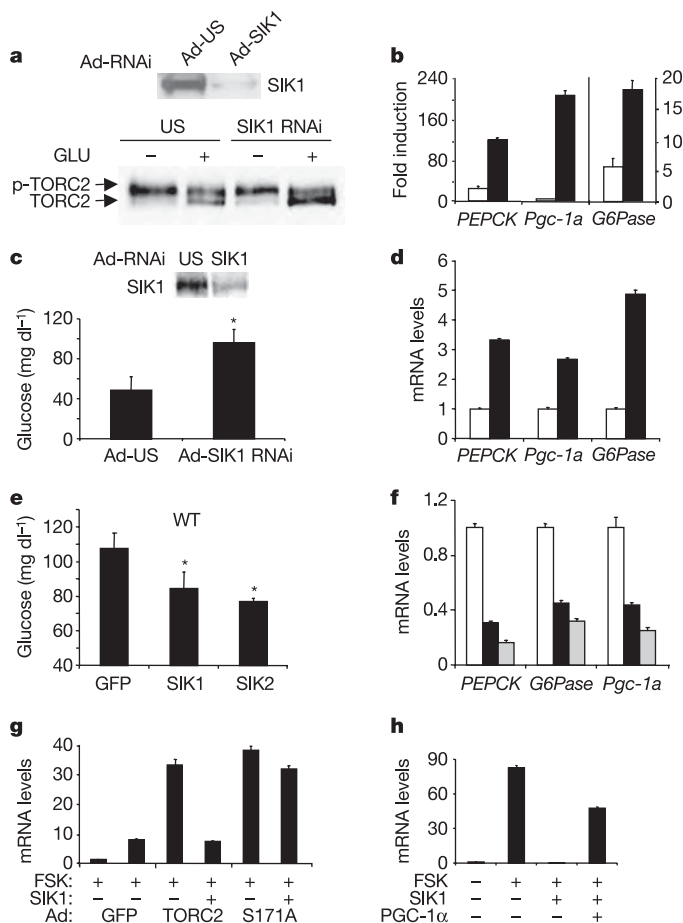


Figure 4 | SIKs inhibit hepatic gluconeogenesis via Ser 171 phosphorylation of TORC2. **a, b**, Effect of Ad-US (open columns) and Ad-SIK1 RNAi (filled columns) on SIK1, TORC2 and p-TORC2 protein levels (**a**) and on mRNA levels for gluconeogenic genes (**b**) in primary hepatocytes exposed to glucagon. **c, d**, Fasting glucose (**c**) and gluconeogenic mRNA levels (**d**) in Ad-US (open columns in **d**) or Ad-SIK1 RNAi (filled columns in **d**) mice fasted for 4 h (asterisk, $P < 0.01$, t -test). Inset in **c** shows a western blot of hepatic SIK1 levels in Ad-US and Ad-SIK1 RNAi mice. **e, f**, Effect of Ad-SIK1 (black columns in **f**), Ad-SIK2 (grey columns in **f**) or control Ad-GFP (white columns in **f**) on fasting glucose levels (**e**) (asterisk, $P < 0.01$, t -test) ($n = 4$) and on mRNA levels for gluconeogenic genes (**f**). **g, h**, Effect of Ad-SIK1 on *PEPCK* mRNA levels in primary hepatocytes co-expressing Ad-TORC2 and Ad-TORC2(Ser171Ala) (**g**) or Ad-PGC-1 α (**h**). Treatment with FSK is as indicated. Data in **b–h** represent mean $\pm$ s.e.m.

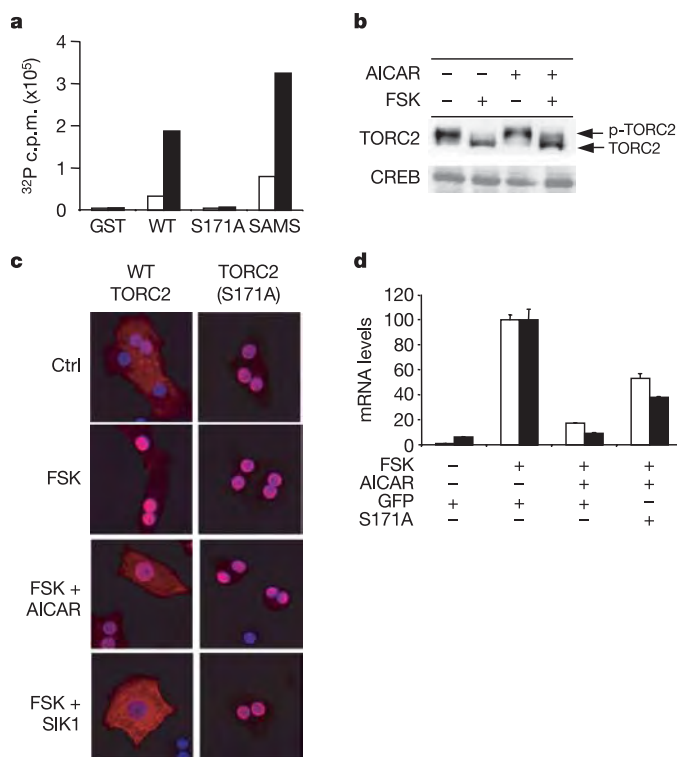


Figure 5 | AMPK inhibits activation of TORC2 by fasting signals. **a**, *In vitro* phosphorylation of wild-type, GST–TORC2(Ser171Ala) (161–181), GST, or optimal substrate (SAMS) by activated AMPK with (filled columns) or without (open columns) AMP addition to reactions. c.p.m., counts per min. **b**, Effect of AMPK inducer AICAR on p-TORC2 levels in primary hepatocytes. Treatment with FSK is indicated. **c**, Immunocytochemical analysis showing effects of AICAR and Ad-SIK1 on Ad-TORC2 and Ad-TORC2(Ser171Ala) localization in primary hepatocytes. **d**, Effect of AICAR on expression of gluconeogenic genes in primary hepatocytes. *PEPCK*, open columns; *Pgc-1 α* , filled columns. Rescue with Ad-TORC2(Ser171Ala) is indicated. Data represent mean $\pm$ s.e.m.

control mice (Supplementary Fig. S12). Ad-SIK1 and Ad-SIK2 were also effective in reducing blood glucose levels in fasted *db/db* diabetic mice, which exhibit hepatic insulin resistance (Supplementary Fig. S13). Indeed, SIK expression did not improve insulin signalling in *db/db* animals, as measured by western blot analysis of phosphorylated (Ser 473) Akt levels in liver extracts from mice after acute intraperitoneal insulin injection (10 min; Supplementary Fig. S14). Taken together these results suggest that SIKs disrupt hepatic gluconeogenesis via an insulin-independent mechanism.

To evaluate whether SIK1 inhibits the gluconeogenic programme via TORC2 phosphorylation, we prepared an adenovirus expressing phosphorylation-defective TORC2 (TORC2(Ser171Ala)). Wild-type and mutant Ad-TORC2(Ser171Ala) were expressed at comparable levels and had similar effects on *PEPCK* and *PGC-1 α* gene expression in primary hepatocytes exposed to FSK (Fig. 4g; see also Supplementary Fig. S15). Ad-SIK1 blocked wild-type TORC2 activity almost completely but had little effect on TORC2(Ser171Ala), demonstrating the importance of Ser 171 for inhibition of the gluconeogenic programme by this kinase (Fig. 4g). Indeed, Ad-PGC-1 α was competent to rescue *PEPCK* expression in Ad-SIK1 infected cells, confirming the notion that SIK1 inhibits hepatic gene expression specifically at the level of TORC2 phosphorylation (Fig. 4h).

AMPK blocks TORC2 activation

Activation of the AMPK pathway in response to energy stress has been shown to block expression of the gluconeogenic programme, although the relevant targets for inhibition have not been identified^{14,25}. Changes in cellular ATP levels seem to be uniquely sensed by AMPK, the founding member of this family^{26–28}. The presence of a consensus AMPK phosphorylation site (Ser 171) that modulates TORC2 activity in liver prompted us to consider whether, in a manner analogous to SIK1, AMPK triggers TORC2 phosphorylation in response to ATP depletion and thereby inhibits hepatic gluconeogenesis. Activated AMPK was found to phosphorylate a wild-type but not Ser171Ala mutant GST-TORC2 (161–181) peptide *in vitro* at levels comparable to an optimal AMPK peptide substrate (SAMS)²⁷ (Fig. 5a). Addition of AMP to reactions further stimulated Ser 171 phosphorylation by AMPK on both GST-TORC2(Ser171Ala) (161–181) and GST-TORC2 (1–240) polypeptides (Supplementary Fig. S16). Consistent with this result, selective activation of cellular AMPK by exposure of primary hepatocytes to the AMP analogue 5-aminoimidazole-4-carboxamide riboside (AICAR) triggered phosphorylation of endogenous TORC2 twofold even in the presence of FSK (Fig. 5b; see also Supplementary Figs S17 and S18).

Given that it lies on a different signalling axis from the cAMP–PKA pathway, AMPK would be predicted to inhibit nuclear entry of TORC2 in cells exposed to FSK. Using primary rat hepatocytes infected with Ad-TORC2 constructs, we found that FSK triggered translocation of wild-type TORC2 (Fig. 5c). Confirming the importance of Ser 171 in this regard, mutant TORC2(Ser171Ala) protein remained constitutively nuclear under both conditions. Treatment with AICAR inhibited nuclear entry of wild-type but not TORC2(Ser171Ala) in cells exposed to FSK; Ad-SIK1 similarly blocked TORC2 nuclear entry in a Ser 171-dependent manner (Fig. 5c).

The ability of AICAR to limit TORC2 nuclear entry in response to fasting signals suggests that AMPK disrupts gluconeogenesis via Ser 171 phosphorylation. Indeed, exposure of primary hepatocytes to AICAR blocked induction of *PEPCK* and *PGC-1 α* genes in response to FSK (Fig. 5d). However, phosphorylation-defective Ad-TORC2(Ser171Ala) restored expression of *G6Pase*, *PEPCK* and *PGC-1 α* in the presence of AICAR inhibitor, demonstrating the importance of Ser 171 in mediating inhibitory effects of AMPK on gluconeogenic genes (Fig. 5d; see also Supplementary Fig. S19).

Discussion

Our studies suggest that TORC2 is a critical switch that modulates the gluconeogenic programme in response to both hormonal and

intracellular signals. Although the CREB–CBP interaction itself did not seem to be critical for hepatic gene activation during fasting, we imagine that CBP and P300 together provide necessary histone acetylase activities, which further enhance TORC2 function in this setting.

TORC2 activity is tightly regulated during fasting by SIK1, which forms part of an auto-regulatory loop that attenuates the gluconeogenic programme by phosphorylating TORC2 at Ser 171 and thus promoting its export to the cytoplasm. Other SIK family members (SIK2 and SIK3) probably inhibit TORC2 activity as well, particularly under feeding conditions when SIK1 levels are low.

Superimposed on this regulatory circuit, TORC2 activity is additionally modulated at the cellular level by AMPK, the induction of which disrupts hepatic glucose production in response to stressors. Indeed, the ability of AMPK to catalyse Ser 171 phosphorylation even in the presence of cAMP agonist suggests that this pathway may override the inductive effects of hormones such as glucagon on fasting metabolism. A number of adipokines such as adiponectin¹⁴ and resistin¹³ have been shown to modulate hepatic gluconeogenesis via an AMPK-dependent mechanism; our results provide a regulatory framework with which to test these regulatory inputs. Metformin, a compound that also activates AMPK²⁹, has been widely used for treatment of type 2 diabetes owing to its effects in blocking hepatic gluconeogenesis and in stimulating glucose uptake in muscle³⁰. Other compounds that enhance TORC2 phosphorylation in liver may provide similar therapeutic benefit for individuals with insulin resistance.

METHODS

Primary cultures. Primary rat hepatocytes were prepared from Sprague–Dawley rats as described⁹. After 16-h adenoviral infection, cells were maintained in serum-free medium 199 (Invitrogen) for 24 (cDNA expression adenoviruses) or 48 h (RNAi adenoviruses); cells were exposed to 10 μ M forskolin or 100 nM glucagon for 2 h unless noted otherwise. Cells were treated with AICAR (1 mM) for either 30 min (western blot, immunostaining) or 2 h (quantitative PCR analysis). CHX (10 μ g ml^{–1}) was added 30 min before addition of 100 nM glucagon. Glucose output assays were performed as reported using a glucose Trinder kit (Sigma)³¹. Cells infected with TORC1, TORC2 or GFP adenovirus were treated with FSK plus dexamethasone for 4 h or with insulin (100 nM) for 16 h.

Adenoviruses. Control GFP and unspecific RNAi adenoviruses were as described⁹. Adenoviruses for SIK1, SIK2 as well as wild-type and TORC2(Ser171Ala) expression were generated as described⁹.

Animals. Male 7-week-old C57BL6 mice (Harlan) or diabetic *db/db* mice (Jackson laboratory) were used. Adenovirus injections were performed as reported⁹. Sixteen-hour fasting blood glucose levels were measured either 5 days (Ad-TORC2 RNAi, Ad-SIK1/2) or 9 days (Ad-TORC2) after injection. Four-hour fasting blood glucose levels were measured for Ad-SIK1 RNAi mice 2 days after injection. Ad-TORC2 mice (Fig. 1) were injected intraperitoneally with glucagon (5 μ g per kg body weight), insulin (1 unit per kg), or PBS, and livers were collected for immunostaining or western blot analysis after 10 min. Absence of adenovirus-induced hepatitis was confirmed by measuring circulating ALT/AST levels.

Metabolites. Blood glucose levels were monitored from tail vein blood using an automatic glucose monitor (One Touch, Lifescan). Plasma insulin levels were determined using commercial insulin assay kit (ALPCO Diagnostics).

Quantitative PCR. Total RNA was prepared and mRNA levels were measured as previously described⁹.

Western blot. Whole-cell extracts were prepared using SDS-urea-lysis buffer. Western blot assays on 50–100 μ g of protein were performed as described^{19,31,32}. For Fig. 1b, western blot assays were performed on haemagglutinin (HA)-tagged TORC2 immunoprecipitates using phospho (Ser 171) specific and non-discriminating TORC2 antisera. Immunoprecipitation experiments were performed as described¹⁹.

Immunostaining. Formalin-fixed, paraffin-embedded liver sections (5 μ m) were deparaffinized in two changes of xylenes and hydrated by extensive washes with ethanol and distilled H₂O. After antibody incubation (anti-TORC2 (1:1,600), anti-p-CREB (5322, 1:500)), slides were incubated with donkey anti-rabbit Cy3 (1:600; 45 min, room temperature). For primary hepatocytes, cells were fixed in 4% paraformaldehyde for 10 min; anti-HA antibody (1:1,000, Santa Cruz) was used for HA-TORC2 detection. Slides were washed and

mounted with coverslips using Vectashield mounting media containing 4,6-diamidino-2-phenylindole (DAPI).

Chromatin immunoprecipitation. Mice were injected intraperitoneally with either glucagon or insulin for 10 min. Nuclear isolation, cross-linking and chromatin immunoprecipitation assays on liver samples were performed as described¹⁹. Precipitated DNA fragments were analysed by quantitative PCR using primers against relevant mouse promoters.

In vitro kinase assays. Recombinant GST–TORC2 (1–240) and GST–TORC2 (161–181) proteins (wild-type and Ser171Ala) were phosphorylated *in vitro* with 100 mU purified AMPK (Upstate Biotech) with or without 300 μ M AMP according to the manufacturer's instructions. After 15 min incubation with [γ -³²P]ATP, ³²P incorporation was measured by scintillation.

Received 14 April; accepted 27 June 2005.

Published online 7 September 2005.

1. Saltiel, A. & Kahn, C. R. Insulin signalling and the regulation of glucose and lipid metabolism. *Nature* **414**, 799–806 (2001).
2. Saltiel, A. R. New perspectives into the molecular pathogenesis and treatment of type 2 diabetes. *Cell* **104**, 517–529 (2001).
3. Phillips, C. A. & Molitch, M. E. The relationship between glucose control and the development and progression of diabetic nephropathy. *Curr. Diab. Rep.* **2**, 523–529 (2002).
4. Hanson, R. W. & Reshef, L. Regulation of phosphoenolpyruvate carboxykinase (GTP) gene expression. *Annu. Rev. Biochem.* **66**, 581–611 (1997).
5. Hall, R. K. & Granner, D. K. Insulin regulates expression of metabolic genes through divergent signalling pathways. *J. Basic Clin. Physiol. Pharmacol.* **10**, 119–133 (1999).
6. Herzig, S. *et al.* CREB regulates hepatic gluconeogenesis via the co-activator PGC-1. *Nature* **413**, 179–183 (2001).
7. Herzig, S. *et al.* CREB controls hepatic lipid metabolism through nuclear hormone receptor PPAR- γ . *Nature* **426**, 190–193 (2003).
8. Yoon, J. *et al.* Control of hepatic gluconeogenesis through the transcriptional coactivator PGC-1. *Nature* **413**, 131–138 (2001).
9. Koo, S. H. *et al.* PGC-1 promotes insulin resistance in liver through PPAR- α -dependent induction of TRB-3. *Nature Med.* **10**, 530–534 (2004).
10. Lin, J. *et al.* Defects in adaptive energy metabolism with CNS-linked hyperactivity in PGC-1 α null mice. *Cell* **119**, 121–135 (2004).
11. Leone, T. C. *et al.* PGC-1 α deficiency causes multi-system energy metabolic derangements: muscle dysfunction, abnormal weight control and hepatic steatosis. *PLoS Biol.* **3**, e101 (2005).
12. Puigserver, P. *et al.* Insulin-regulated hepatic gluconeogenesis through FOXO1–PGC-1 α interaction. *Nature* **423**, 550–555 (2003).
13. Banerjee, R. R. *et al.* Regulation of fasted blood glucose by resistin. *Science* **303**, 1195–1198 (2004).
14. Yamauchi, T. *et al.* Adiponectin stimulates glucose utilization and fatty-acid oxidation by activating AMP-activated protein kinase. *Nature Med.* **8**, 1288–1295 (2002).
15. Chirivia, J. C. *et al.* Phosphorylated CREB binds specifically to the nuclear protein CBP. *Nature* **365**, 855–859 (1993).
16. Arias, J. *et al.* Activation of cAMP and mitogen responsive genes relies on a common nuclear factor. *Nature* **370**, 226–228 (1994).
17. Conkright, M. D. *et al.* TORCs: transducers of regulated CREB activity. *Mol. Cell* **12**, 413–423 (2003).
18. Iourgenko, V. *et al.* Identification of a family of cAMP response element-binding protein coactivators by genome-scale functional analysis in mammalian cells. *Proc. Natl Acad. Sci. USA* **100**, 12147–12152 (2003).
19. Screaton, R. A. *et al.* The CREB coactivator TORC2 functions as a calcium- and cAMP-sensitive coincidence detector. *Cell* **119**, 61–74 (2004).
20. Bittinger, M. A. *et al.* Activation of cAMP response element-mediated gene expression by regulated nuclear transport of TORC proteins. *Curr. Biol.* **14**, 2156–2161 (2004).
21. Zhou, X. Y. *et al.* Insulin regulation of hepatic gluconeogenesis through phosphorylation of CREB-binding protein. *Nature Med.* **10**, 633–637 (2004).
22. Kasper, L. H. *et al.* A transcription-factor-binding surface of coactivator p300 is required for haematopoiesis. *Nature* **419**, 738–743 (2002).
23. Ahn, S. *et al.* A dominant negative inhibitor of CREB reveals that it is a general mediator stimulus-dependent transcription of c-fos. *Mol. Cell. Biol.* **18**, 967–977 (1998).
24. Sasaki, K. *et al.* Multihormonal regulation of phosphoenolpyruvate carboxykinase gene transcription. *J. Biol. Chem.* **259**, 15242–15251 (1984).
25. Lochhead, P. A., Coghlan, M., Rice, S. Q. & Sutherland, C. Inhibition of GSK-3 selectively reduces glucose-6-phosphatase and phosphatase and phosphoenolpyruvate carboxykinase gene expression. *Diabetes* **50**, 937–946 (2001).
26. Kahn, B. B., Alquier, T., Carling, D. & Hardie, D. G. AMP-activated protein kinase: Ancient energy gauge provides clues to modern understanding of metabolism. *Cell Metab.* **1**, 15–25 (2005).
27. Lizcano, J. M. *et al.* LKB1 is a master kinase that activates 13 kinases of the AMPK subfamily, including MARK/PAR-1. *EMBO J.* **23**, 833–843 (2004).
28. Sakamoto, K., Goransson, O., Hardie, D. G. & Alessi, D. R. Activity of LKB1 and AMPK-related kinases in skeletal muscle: effects of contraction, phenformin, and AICAR. *Am. J. Physiol. Endocrinol. Metab.* **287**, E310–E317 (2004).
29. Radziuk, J., Bailey, C. J., Wiernsperger, N. F. & Yudkin, J. S. Metformin and its liver targets in the treatment of type 2 diabetes. *Curr. Drug Targets Immune Endocr. Metab. Disord.* **3**, 151–169 (2003).
30. Bergeron, R. *et al.* Effect of 5-aminoimidazole-4-carboxamide-1- β -D-ribofuranoside infusion on *in vivo* glucose and lipid metabolism in lean and obese Zucker rats. *Diabetes* **50**, 1076–1082 (2001).
31. Du, K., Herzig, S., Kulkarni, R. N. & Montminy, M. TRB3: a tribbles homolog that inhibits Akt/PKB activation by insulin in liver. *Science* **300**, 1574–1577 (2003).
32. Katoh, Y. *et al.* Salt-inducible kinase-1 represses cAMP response element-binding protein activity both in the nucleus and in the cytoplasm. *Eur. J. Biochem.* **271**, 4307–4319 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by grants from the NIH, the Keckhefer Foundation, the Hillblom Foundation, the American Diabetes Association, the Juvenile Diabetes Foundation and the Leukemia and Lymphoma Society. We thank L. Vera for mouse injections and ProteinExpress Co. Ltd for the gift of p-TORC2 antiserum.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.M. (montminy@salk.edu).

Rebuilt AAA+ motors reveal operating principles for ATP-fuelled machines

Andreas Martin¹, Tania A. Baker^{1,2} & Robert T. Sauer¹

Hexameric ring-shaped ATPases of the AAA+ (for ATPases associated with various cellular activities) superfamily power cellular processes in which macromolecular structures and complexes are dismantled or denatured, but the mechanisms used by these machine-like enzymes are poorly understood. By covalently linking active and inactive subunits of the ATPase ClpX to form hexamers, here we show that diverse geometric arrangements can support the enzymatic unfolding of protein substrates and translocation of the denatured polypeptide into the ClpP peptidase for degradation. These studies indicate that the ClpX power stroke is generated by ATP hydrolysis in a single subunit, rule out concerted and strict sequential ATP hydrolysis models, and provide evidence for a probabilistic sequence of nucleotide hydrolysis. This mechanism would allow any ClpX subunit in contact with a translocating polypeptide to hydrolyse ATP to drive substrate spooling into ClpP, and would prevent stalling if one subunit failed to bind or hydrolyse ATP. Energy-dependent machines with highly diverse quaternary architectures and molecular functions could operate by similar asymmetric mechanisms.

AAA+ molecular machines power the degradation of proteins, remodelling and disassembly of macromolecular complexes, solubilization of aggregates, and translocation of proteins and nucleic acids^{1–7}. Energy-dependent protein degradation is carried out by the proteasome in eukaryotes, and by the proteases ClpXP, ClpAP, HslUV, Lon, FtsH and other related enzymes in bacteria, mitochondria and chloroplasts. In these proteases, AAA+ ATPases recognize appropriate substrates, denature these proteins, and translocate the unfolded polypeptide through a central pore into the proteolytic chamber of an associated peptidase.

ClpX is a powerful protein-unfolding machine with six identical subunits, each containing one AAA+ ATPase domain⁶. Hexameric rings of ClpX assemble with the ClpP peptidase to form the ClpXP protease⁸. ATP-bound ClpX binds to exposed peptide signals on protein substrates. ClpX then catalyses substrate denaturation and translocation of the unfolded polypeptide into ClpP in a process that can require hundreds of cycles of ATP hydrolysis^{9–14}. Therefore, determining how ATP hydrolysis is coupled to these mechanical tasks is critical for understanding how this molecular machine operates. Each ClpX subunit has the same sequence, making models involving ATP binding and hydrolysis in an all-or-none fashion appealing. However, recent evidence indicates that individual subunits have distinct functional roles, as several subunits are nucleotide-free in ATP-saturated ClpX hexamers, and nucleotide-bound subunits exhibit a range of different properties¹⁵. Experiments with other hexameric ATPases, including helicases and translocases, also demonstrate multiple classes of nucleotide-binding sites^{16–22}. Such behaviour has generally been interpreted as evidence for ‘binding-change’ mechanisms, in which the conformation, nucleotide state and catalytic properties of individual subunits change in a rotary sequence, as initially proposed for the F₁ ATPase (ref. 23). Nevertheless, AAA+ and related enzymes adopt a puzzling assortment of symmetric and asymmetric quaternary forms, suggesting that these machines may work in different ways.

Covalent linkage of ClpX subunits

To study the functional roles of individual subunits of *Escherichia coli* ClpX, we engineered genes encoding covalently linked variants of the ATPase. Many flexible linkers were used to connect full-length ClpX subunits, but the resulting proteins were insoluble. We then used a truncated form of ClpX missing the amino-terminal domain (ClpX-ΔN), which still powers ClpP-mediated degradation of proteins bearing the *ssrA* degradation tag²⁴. ClpX-ΔN subunits remained soluble when linked with the sequence ASGAGGSEGGGSEGGTSGAT (designated L20). Modelling based on the *Helicobacter pylori* enzyme²⁵ showed that L20 could connect the C terminus of one subunit to the N terminus of the clockwise subunit (as shown in the top view of Fig. 1a), but was too short to connect to the N terminus of the counter-clockwise subunit or to non-adjacent subunits. Specifically, L20 spans ~70 Å when fully extended; the clockwise connection required ~45 Å, whereas the counter-clockwise connection required ~95 Å.

We initially constructed a ClpX variant containing two linked ClpX-ΔN subunits, and subsequently built molecules with three and six linked subunits. Figure 1b shows the gene encoding the six-subunit protein. During the purification procedure, gel filtration showed that each protein chromatographed as a pseudo-hexamer (a trimer of two modules, a dimer of three modules, and so on) with the same radius of gyration as the unlinked hexamer (Fig. 1c). Notably, the purified linked ClpX-ΔN proteins (Fig. 1d) were comparable to wild-type ClpX in supporting ClpP-mediated degradation of an *ssrA*-tagged substrate and in ATP hydrolysis (Table 1). Hence, covalent connection of ClpX-ΔN subunits results in active pseudo-hexameric enzymes.

Two previously characterized mutations were used to eliminate ATP hydrolysis in individual ClpX subunits^{15,26}. The E185Q mutation in the Walker B motif of ClpX prevents ATP hydrolysis, but allows ClpX to change conformation in response to ATP binding and thus bind ClpP and the *ssrA* tag of substrates¹⁵. The R370K mutation in the sensor-2 motif also blocks hydrolysis, but uncouples

¹Department of Biology, ²Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

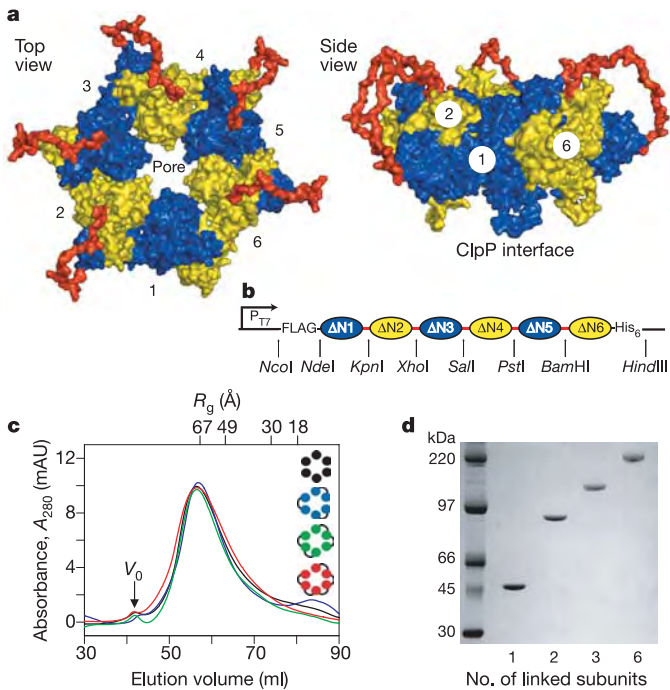


Figure 1 | Covalent connection of ClpX subunits to form single-chain pseudo-hexamers. **a**, Models of the L20 linker (red) connecting six ClpX- Δ N subunits (alternating yellow and blue) based on the crystal structure of *H. pylori* ClpX- Δ N (residues 75–438; ref. 25). **b**, Gene encoding six linked subunits of *E. coli* ClpX- Δ N (residues 61–423) with FLAG and His₆ epitopes. **c**, WW/WW/WW (blue), WWW/WWW (green) and WWWWWWW (red) eluted as pseudo-hexamers (radius of gyration $\sim$ 71 Å) during gel filtration. Based on the *H. pylori* structure²⁵, the longest dimension of the *E. coli* ClpX- Δ N hexamer (black) is 70 Å. AU, absorbance units. **d**, SDS–polyacrylamide gel electrophoresis of purified enzymes after gel filtration.

ATP binding from the conformational changes required for ClpP and ssrA-tag binding²⁶, thus trapping ClpX subunits in an inactive ‘resting’ conformation. We refer to the subunits as W (for wild type), E (for E185Q) or R (for R370K), and to the linked modules using names such as WR or RWE. Variants composed of EE or RR modules formed catalytically inert pseudo-hexamers (designated EE/EE/EE and RR/RR/RR) with the same binding properties as the unlinked E and R ClpX mutants^{15,26} (data not shown).

Symmetric arrangements of active and inactive subunits

Some properties of ClpX and related ATPases suggest that the hexamers assemble and function with trimer-of-dimers symmetry, whereas other experiments suggest a dimer-of-trimers symmetry^{19,20}. The results presented below show that ClpX- Δ N variants with mixtures of subunits configured in either symmetry are functional, including some arrangements containing just two active subunits.

A trimer-of-dimers variant (WR/WR/WR), with alternating active and resting subunits in the hexamer, mediated ClpP-dependent degradation of the denatured substrate carboxymethyl (CM)–titin–ssrA (ref. 13) at $\sim$ 35% of the rate of the WW/WW/WW parental protein, and was $\sim$ 45% as active as the wild-type complex in ATP hydrolysis during protein degradation (Fig. 2a; Table 1). The ATPase activity per wild-type subunit of WR/WR/WR was close to that of the parental enzyme (Table 1). This result is clearly inconsistent with models in which all six ClpX subunits must bind and hydrolyse ATP in a concerted fashion.

We next built a dimer-of-trimers variant (WWR/WWR) from modules containing two active subunits and one resting subunit. This mutant showed $\sim$ 60% of wild-type activity in protein degradation and ATPase assays (Table 1). Again, activity was roughly proportional to the number of active subunits. The activities of WRW/WRW and RWW/RWW variants were found to be within error of WWR/WWR (data not shown). Because these three variants are related by circular permutation, the interfaces between covalently linked and non-covalently linked subunits appear to be functionally equivalent.

To investigate enzymes containing two active subunits located opposite each other in the hexamer, we constructed RWE/RWE, WRE/WRE and WEE/WEE variants. Each mutant had at least four

Table 1 | ATPase and protein degradation activities of linked variants of ClpX

	CM-titin-ssrA degradation					Native substrate degradation		
	ATPase per enzyme* (min ⁻¹)	ATPase per wild-type subunit† (min ⁻¹)	V _{max} per enzyme‡ (min ⁻¹)	V _{max} per wild-type subunit§ (min ⁻¹)	K _m ‡ (μM)	V _{max} per enzyme titin-ssrA (min ⁻¹)	V _{max} per enzyme V15P titin-ssrA (min ⁻¹)	V _{max} per enzyme Y9P titin-ssrA (min ⁻¹)
Wild-type ClpX	430	72	4.12	0.68	1.2	0.25	0.85	1.50
ClpX- Δ N variant								
W/W/W/W/W/W	421	70	–	–	–	–	–	–
WW/WW/WW	415	69	3.95	0.66	2.1	–	–	–
WWW/WWW	412	69	3.94	0.66	2.0	0.24	0.81	1.11
WWWWW	449	75	3.81	0.64	1.6	–	–	–
WR/WR/WR	189	63	1.26	0.42	2.2	–	–	–
WWR/WWR	250	63	2.39	0.60	1.8	0.14	0.52	0.52
WWE/WWE	95	24	1.12	0.28	1.6	–	–	–
RWE/RWE	120	60	1.14	0.57	1.5	0.07	0.14	0.30
WRE/WRE	41	21	0.14	0.07	–	–	–	–
WEE/WEE	31	16	0.10	0.05	–	–	–	–
WWWWR	277	55	2.86	0.56	1.4	0.21	0.61	0.84
WWWWR	247	62	2.32	0.55	2.2	0.11	0.36	0.47
WWWRR	287	96	1.81	0.57	2.2	0.07	0.27	0.33
WWREER	101	51	0.21	0.11	1.6	–	–	–
WWEEER	51	26	0.11	0.06	–	–	–	–
WEREER	79	79	0.14	0.14	1.8	0.0049	0.014	0.030

*ATPase activity of the hexamer, based on replicate measurements ($n = 5$). ATPase rates (per enzyme min⁻¹) have an error of $\pm 5\%$.
†ATPase activity divided by the number of wild-type subunits.
‡The errors in V_{max} and K_m were estimated to be $\pm 10\%$ on the basis of nonlinear least-squares fitting of Michaelis–Menten plots. Replicate measurements ($n = 3$) of velocity at the highest substrate concentration tested were within 10% of V_{max} .
§ V_{max} divided by the number of wild-type subunits.
||All linked variants chromatographed as pseudo-hexamers on gel filtration.

W and E subunits to maintain ClpP and *ssrA*-tag binding¹⁵ (WRR/WRR did not bind ClpP or *ssrA* peptide; data not shown). RWE/RWE was ~30% as active as WWW/WWW in ATP hydrolysis and ClpP-dependent degradation of CM-titin-*ssrA* (Fig. 2a; Table 1). Thus, two opposed wild-type subunits can power protein degradation with a velocity proportional to the number of active subunits in the hexamer. In contrast, WRE/WRE and WEE/WEE were only ~3% as active as WWW/WWW in protein degradation (tenfold less active than RWE/RWE) and 7–10% as active as WWW/WWW in ATP hydrolysis (about 3.5-fold less active than RWE/RWE; Fig. 2a; Table 1). Directional communication between the active subunit and its neighbours apparently has a key role in setting enzyme activity levels, with the highest activity observed when a resting subunit is the counter-clockwise neighbour (top view, Fig. 1a).

Asymmetric variants with six linked subunits

Pseudo-hexamers of the variants discussed above assemble as symmetric oligomers of two or three linked subunits. To investigate asymmetric combinations of active and inactive subunits, we constructed molecules containing six linked subunits. The WWWWR, WWWRR and WWWRRR variants showed ATPase and protein degradation activities proportional to the number of wild-type subunits (Table 1). Hence, a symmetric arrangement of active subunits is not required for ClpX function.

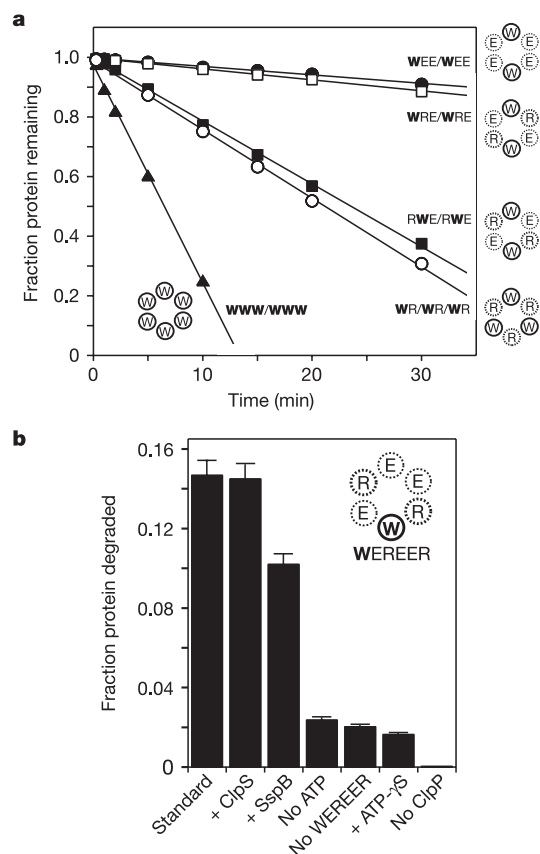


Figure 2 | ClpX hexamers with 1–3 active subunits drive ClpP-mediated degradation of an unfolded substrate. **a**, Kinetics of degradation of denatured CM-titin-*ssrA* (15 μ M) by ClpP and WEE/WEE, WRE/WRE, RWE/RWE, WR/WR/WR and WWW/WWW. **b**, WEREER-mediated degradation of CM-titin-*ssrA* (15 μ M), measured after 60 min, required ClpP (0.9 μ M), ATP (5 mM) and WEREER (0.3 μ M) (standard reaction). Degradation was inhibited slightly by SspB (15 μ M), as expected for a ClpX- Δ N variant but not for wild-type ClpXP (ref. 28), but was not affected by ClpS (1 μ M), which inhibits degradation of *ssrA*-tagged substrates by ClpAP (ref. 27). Error bars represent one standard deviation ($n = 3$).

To determine whether pseudo-hexamers with two adjacent wild-type subunits or a single active subunit could function, we built WWREER, WWEEER and WEREER variants. These mutants showed low levels of activity in ClpP-mediated degradation of CM-titin-*ssrA* (Table 1). Because WEREER possessed only ~3% of the activity of WWWWWW, controls were performed to test for a contaminating protease. Full enzymatic activity required ATP and ClpP in addition to WEREER (Fig. 2b). Moreover, activity was neither inhibited by ClpS, which blocks ClpAP-mediated degradation of *ssrA*-tagged substrates²⁷, nor stimulated by SspB, which enhances degradation of *ssrA*-tagged substrates by wild-type ClpXP (ref. 28) (Fig. 2b). In fact, SspB inhibited protein degradation slightly, as expected for a ClpX- Δ N variant⁶. Thus, WEREER itself, and not a contaminant, is able to drive protein degradation. Conformational changes resulting from ATP binding and hydrolysis in a single subunit in the hexamer must represent the basic power stroke of ClpX and are sufficient to drive ClpP-dependent protein degradation.

Slow degradation is caused by slow translocation

Initial rates of CM-titin-*ssrA* proteolysis for nine mutant enzymes and three parental enzymes were measured across a series of substrate concentrations and fitted to the Michaelis-Menten equation to determine the K_m and V_{max} (Table 1). The K_m values for all enzymes were similar; mutant V_{max} values varied from 3–80% of the wild-type value. Thus, the mutants bind CM-titin-*ssrA* normally, but are less active because of slower catalysis. In principle, these lower V_{max} values could reflect slower translocation of the substrate from ClpX to ClpP, or slower initiation of translocation. For the RWE/RWE enzyme, we tested these alternatives by measuring V_{max} for substrates containing repeats of the unfolded CM-titin domain²⁹. As expected for a model in which the rate of substrate translocation determines the maximal rate of ClpP degradation, the average time required to degrade CM-titin substrates with one, two or three repeats varied linearly with substrate length and gave a y -intercept close to zero (Fig. 3a). The other mutant variants are presumably also less active because of slower translocation.

Energetic efficiency of degradation

Energy use by different linked enzymes was calculated from the rates of ATP hydrolysis and CM-titin-*ssrA* degradation (Fig. 3c). For WWWWWW, the cost of degradation was roughly 100 ATP molecules per molecule of substrate. ATP consumption was three- to fivefold higher for five variants, each with two or fewer active subunits. For these enzymes, translocation may be inefficient because a second active subunit is not present, is not in the correct state, or is not suitably positioned to coordinate effective transport or to prevent slipping. However, for seven mutants the cost of degradation was very similar to the wild-type enzyme (Fig. 3b). This result demonstrates that comparable thermodynamic efficiency can be achieved with very different arrangements of active and inactive subunits. For example, RWE/RWE/ClpP degrades CM-titin-*ssrA* with the same energetic efficiency as the wild-type complex despite having just two active ClpX- Δ N subunits.

Unfolding stable native substrates

ClpXP denatures native proteins before degradation^{9–11,13,14,30}. To determine whether linked mutants unfold stable native substrates, we assayed the degradation of titin-*ssrA* and its variants with the destabilizing Y9P and V15P mutations (Y9P titin-*ssrA* and V15P titin-*ssrA*). Enzymatic denaturation is the rate-limiting step in ClpXP proteolysis of these proteins, and stability differences result in degradation rates in the order Y9P titin-*ssrA* > V15P titin-*ssrA* > titin-*ssrA* (ref. 13). The WEREER, RWE/RWE, WWWRRR, WWR/WR, WWWRR and WWWWR variants powered ClpP-mediated degradation of each native substrate (Fig. 3c). WEREER/ClpP was the least active enzyme, but still degraded

© 2005 Nature Publishing Group

hydrolysis in a specific subunit could depend on nucleotide binding in adjacent subunits, structural constraints or substrate interactions. For example, any subunit of ClpX in contact with a translocating substrate might be stimulated to hydrolyse nucleotide and drive the next translocation step, explaining why ClpX hydrolyses ATP more rapidly when it is engaged in polypeptide translocation¹³. We note that the sixfold symmetry of the ClpX hexamer bears no simple relationship to potential conformations of translocating polypeptides, which are highly divergent in sequence and thus in physical properties⁶. As a consequence, one subunit in the ClpX ring might be better positioned than another to interact with irregular features of any given peptide sequence, making a probabilistic model highly attractive. This model would also allow ClpX to avoid stalling if one or more subunits were transiently unable to hydrolyse ATP. In contrast, if the sequence of ATP hydrolysis were preset, then the subunit programmed to fire next might not be in contact with the substrate, leading to stalling or wasted hydrolysis, and inefficient translocation. Probabilistic mechanisms may be especially important for molecular machines that work with highly diverse substrates.

The mechanism we propose for ClpX requires plasticity in nucleotide interactions, as observed in crystal structures of the related HslU ATPase, which show distinct nucleotide-binding geometries¹⁹ (Fig. 4d). It also requires hydrolysis of a single ATP molecule per enzymatic cycle, consistent with studies of ATP-fuelled helicases and DNA translocases^{16,17}. If symmetric interactions with ATP and strict sequential firing are not required, then AAA+ machines and related enzymes, built from a single type of subunit, could function similarly whether they form hexamers^{1–7,19,20,31}, heptamers^{32,33}, or open or closed rings³⁴. Experiments using linked variants should help to distinguish between sequential and probabilistic mechanisms for many of these enzymes. Some ATP-dependent machines (for

example, clamp loader, MCM_{2–7} helicase, dynein and so on) assemble and function using different types of subunits or linked ATPase domains, only some of which are catalytic^{35–37}. The ability of ClpX to function using only one or two catalytically active subunits suggests that AAA+ machines built from a single subunit type could also share common asymmetric mechanisms with these highly diverse molecular machines.

METHODS

Mutant genes were constructed by polymerase chain reaction (PCR) and cloned into pACYCDuet-1 (Novagen). Proteins were expressed at 25 °C in the *recA*[–] *E. coli* strain BLR (DE3) and purified on a Ni²⁺-NTA affinity column (Qiagen) and by chromatography on Sephacryl S-300 (Amersham). Confirming the gel filtration results, WWW/WWW sedimented as a pseudo-hexamer in analytical-equilibrium centrifugation (molecular mass of 246 kDa; expected 250 kDa). Degradation of ³⁵S-labelled titin–ssrA substrates by 0.9 μM ClpP₁₄ and 0.3 μM ClpX-ΔN variant (pseudo-hexamer) was assayed by release of acid-soluble peptides at 30 °C as described¹³. ATPase assays were performed with 0.3 μM ClpX-ΔN pseudo-hexamer, 0.9 μM ClpP₁₄ and 20 μM CM–titin–ssrA at 30 °C as described¹¹.

Normalized protein degradation and ATPase activities (Fig. 3d) were predicted as $a_1WWW + a_2WWE + a_3WWR + a_4EWW + a_5EWE + a_6EWR + a_7RWW + a_8RWE + a_9RWR$. The constants (a_1 – a_9) represent fractional protein degradation or ATPase activities for each class of three-subunit module with a central W subunit. The variables represent the number of such modules in a given variant. Thus, WWWWWW contains 6 WWW modules, RWW/RWW contains 2 RWW and 2 WWR modules, and WWWRRR contains 1 WWW, 1 WWR and 1 RWW module. Modules with a central E or R subunit were assigned zero activity. Values for a_1 – a_9 were calculated by multiple-linear regression (analysis of variance, ANOVA, significance 1.09×10^{-11}) using normalized ATPase and protein degradation rates for 15 ClpX-ΔN variants (30 observations in total; 15 independent observations). For cross-validation, regression was performed using 15 training observations, and R^2 was calculated for the 15 remaining observations. In six trials using different combinations of ATPase and protein degradation rates for the training set, R^2 was 0.91–0.97.

Received 31 May; accepted 14 July 2005.

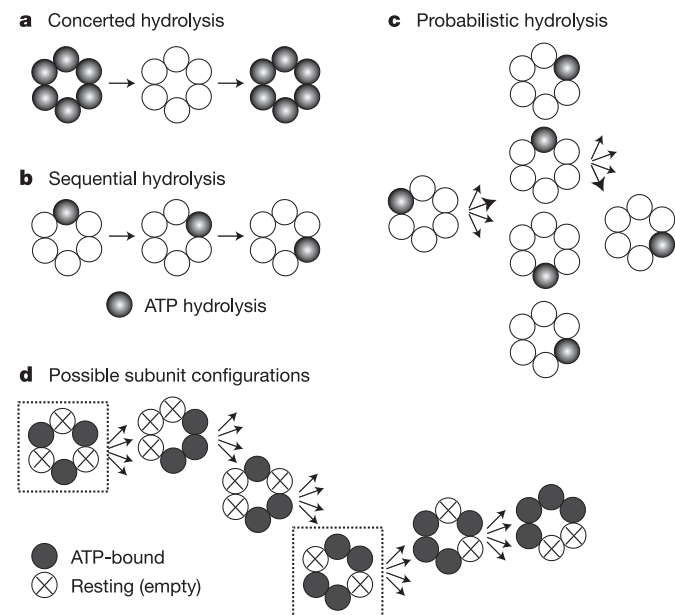


Figure 4 | Models of ATP hydrolysis. **a**, Concerted model. In this model, the simultaneous binding and hydrolysis of six ATP molecules occurs. In variations of this model, fewer than six ATP molecules could undergo synchronous hydrolysis. **b**, Sequential model. ATP hydrolysis in one subunit is followed sequentially by hydrolysis in the neighbouring clockwise subunit, and so on. **c**, Probabilistic model. Sequential trajectories are possible but not obligatory. **d**, If 2–3 ClpX subunits are always nucleotide-free¹⁵, probabilistic ATP binding and hydrolysis could lead to many different configurations of ATP-bound and thus potentially active subunits in the hexamer. The boxed configurations are similar to ones observed in crystal structures of HslU, a ClpX homologue¹⁹.

- Gottesman, S. Proteases and their targets in *Escherichia coli*. *Annu. Rev. Genet.* **30**, 465–506 (1996).
- Neuwald, A. F., Aravind, L., Spouge, J. L. & Koonin, E. V. AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res.* **9**, 27–43 (1999).
- Langer, T. AAA proteases: cellular machines for degrading membrane proteins. *Trends Biochem. Sci.* **25**, 247–251 (2000).
- Ogura, T. & Wilkinson, A. J. AAA+ superfamily ATPases: common structure—diverse function. *Genes Cells* **6**, 575–597 (2001).
- Gottesman, S. Proteolysis in bacterial regulatory circuits. *Annu. Rev. Cell Dev. Biol.* **19**, 565–587 (2003).
- Sauer, R. T. *et al.* Sculpting the proteome with AAA+ proteases and disassembly machines. *Cell* **119**, 9–18 (2004).
- Pickart, C. M. & Cohen, R. E. Proteasomes and their kin: proteases in the machine age. *Nature Rev. Mol. Cell Biol.* **5**, 177–187 (2004).
- Grimaud, R., Kessel, M., Beuron, F., Steven, A. C. & Maurizi, M. R. Enzymatic and structural similarities between the *Escherichia coli* ATP-dependent proteases, ClpXP and ClpAP. *J. Biol. Chem.* **273**, 12476–12481 (1998).
- Kim, Y. I., Burton, R. E., Burton, B. M., Sauer, R. T. & Baker, T. A. Dynamics of substrate denaturation and translocation by the ClpXP degradation machine. *Mol. Cell* **5**, 639–648 (2000).
- Singh, S. K., Grimaud, R., Hoskins, J. R., Wickner, S. & Maurizi, M. R. Unfolding and internalization of proteins by the ATP-dependent proteases ClpXP and ClpAP. *Proc. Natl Acad. Sci. USA* **97**, 8898–8903 (2000).
- Burton, R. E., Siddiqui, S. M., Kim, Y. I., Baker, T. A. & Sauer, R. T. Effects of protein stability and structure on substrate processing by the ClpXP unfolding and degradation machine. *EMBO J.* **20**, 3092–3100 (2001).
- Flynn, J. M. *et al.* Overlapping recognition determinants within the ssrA degradation tag allow modulation of proteolysis. *Proc. Natl Acad. Sci. USA* **98**, 10584–10589 (2001).
- Kenniston, J. A., Baker, T. A., Fernandez, J. M. & Sauer, R. T. Linkage between ATP consumption and mechanical unfolding during the protein processing reactions of a AAA+ degradation machine. *Cell* **114**, 511–520 (2003).
- Bolon, D. N., Grant, R. A., Baker, T. A. & Sauer, R. T. Nucleotide-dependent substrate handoff from the SspB adaptor to the AAA+ ClpXP protease. *Mol. Cell* **16**, 343–350 (2004).
- Hersch, G. L., Burton, R. E., Bolon, D. N., Baker, T. A. & Sauer, R. T. Asymmetric interactions of ATP with the AAA+ ClpX₆ unfoldase: allosteric control of a protein machine. *Cell* **121**, 1017–1027 (2005).

16. Hingorani, M. M., Washington, M. T., Moore, K. C. & Patel, S. S. The dTTPase mechanism of T7 DNA helicase resembles the binding change mechanism of the F₁-ATPase. *Proc. Natl Acad. Sci. USA* **94**, 5012–5017 (1997).
17. Stitt, B. L. & Xu, Y. Sequential hydrolysis of ATP molecules bound in interacting catalytic sites of *Escherichia coli* transcription termination protein Rho. *J. Biol. Chem.* **273**, 26477–26486 (1998).
18. Stitt, B. L. *Escherichia coli* transcription termination factor Rho binds and hydrolyzes ATP using a single class of three sites. *Biochemistry* **40**, 2276–2281 (2001).
19. Bochtler, M. *et al.* The structures of HslU and the ATP-dependent protease HslU–HslV. *Nature* **403**, 800–805 (2000).
20. Singleton, M. R., Sawaya, M. R., Ellenberger, T. & Wigley, D. B. Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides. *Cell* **101**, 589–600 (2000).
21. Zalk, R. & Shoshan-Barmatz, V. ATP-binding sites in brain p97/VCP (valosin-containing protein), a multifunctional AAA ATPase. *Biochem. J.* **374**, 473–480 (2003).
22. Hishida, T., Han, Y. W., Fujimoto, S., Iwasaki, H. & Shinagawa, H. Direct evidence that a conserved arginine in RuvB AAA+ ATPase acts as an allosteric effector for the ATPase activity of the adjacent subunit in a hexamer. *Proc. Natl Acad. Sci. USA* **101**, 9573–9577 (2004).
23. Boyer, P. D. The ATP synthase—a splendid molecular machine. *Annu. Rev. Biochem.* **66**, 717–749 (1997).
24. Wojtyra, U. A., Thibault, G., Tuite, A. & Houry, W. A. The N-terminal zinc binding domain of ClpX is a dimerization domain that modulates the chaperone function. *J. Biol. Chem.* **278**, 48981–48990 (2003).
25. Kim, D. Y. & Kim, K. K. Crystal structure of ClpX molecular chaperone from *Helicobacter pylori*. *J. Biol. Chem.* **278**, 50664–50670 (2003).
26. Joshi, S. A., Hersch, G. L., Baker, T. A. & Sauer, R. T. Communication between ClpX and ClpP during substrate processing and degradation. *Nature Struct. Mol. Biol.* **11**, 404–411 (2004).
27. Dougan, D. A., Reid, B. G., Horwich, A. L. & Bukau, B. ClpS, a substrate modulator of the ClpAP machine. *Mol. Cell* **9**, 673–683 (2002).
28. Levchenko, I., Seidel, M., Sauer, R. T. & Baker, T. A. A specificity-enhancing factor for the ClpXP degradation machine. *Science* **289**, 2354–2356 (2000).
29. Kenniston, J. A., Baker, T. A. & Sauer, R. T. Partitioning between unfolding and release of native domains during ClpXP degradation determines substrate selectivity and partial processing. *Proc. Natl Acad. Sci. USA* **102**, 1390–1395 (2005).
30. Lee, C., Schwartz, M. P., Prakash, S., Iwakura, M. & Matouschek, A. ATP-dependent proteases degrade their substrates by processively unraveling them from the degradation signal. *Mol. Cell* **7**, 627–637 (2001).
31. Gai, D., Zhao, R., Li, D., Finkielstein, C. V. & Chen, X. S. Mechanisms of conformational change for a replicative hexameric helicase of SV40 large tumour antigen. *Cell* **119**, 47–60 (2004).
32. Lee, S. Y. *et al.* Regulation of the transcriptional activator NtrC1: structural studies of the regulatory and AAA+ ATPase domains. *Genes Dev.* **17**, 2552–2563 (2003).
33. Toth, E. A., Li, Y., Sawaya, M. R., Cheng, Y. & Ellenberger, T. The crystal structure of the bifunctional primase-helicase of bacteriophage T7. *Mol. Cell* **12**, 1113–1123 (2003).
34. Skordalakes, E. & Berger, J. M. Structure of Rho transcription terminator: mechanism of mRNA recognition and helicase loading. *Cell* **114**, 135–146 (2003).
35. Schwacha, A. & Bell, S. P. Interactions between two catalytically distinct MCM subgroups are essential for coordinated ATP hydrolysis and DNA replication. *Mol. Cell* **8**, 1093–1104 (2001).
36. Bowman, G. D., Goedken, E. R., Kazmirski, S. L., O'Donnell, M. & Kuriyan, J. DNA polymerase clamp loaders and DNA recognition. *FEBS Lett.* **579**, 863–867 (2005).
37. Sakato, M. & King, S. M. Design and regulation of the AAA+ microtubule motor dynein. *J. Struct. Biol.* **146**, 58–71 (2004).

Acknowledgements We thank S. Bell, D. Bolon, R. Burton, P. Chien, C. Farrell, G. Hersch, J. Kenniston, K. McGinness, S. Moore, F. Solomon and K. Wang for help and discussions. T.A.B. is an employee of the Howard Hughes Medical Institute. This work was supported by grants from the NIH.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.T.S. (bobsauer@mit.edu).

A 'dry' condensation origin for circumstellar carbonates

Alice Toppani^{1,4,†}, François Robert⁴, Guy Libourel^{1,2}, Philippe de Donato³, Odile Barres³, Louis d'Hendecourt⁵ & Jaafar Ghanbaja⁶

The signature of carbonate minerals has long been suspected in the mid-infrared spectra of various astrophysical environments such as protostars¹. Abiogenic carbonates are considered as indicators of aqueous mineral alteration² in the presence of CO₂-rich liquid water. The recent claimed detection of calcite associated with amorphous silicates in two planetary nebulae³ and protostars^{4,5} devoid of planetary bodies questions the relevance of this indicator; but in the absence of an alternative mode of formation under circumstellar conditions, this detection remains controversial^{6–8}. The main dust component observed in circumstellar envelopes is amorphous silicates⁹, which are thought to have formed by non-equilibrium condensation¹⁰. Here we report experiments demonstrating that carbonates can be formed with amorphous silicates during the non-equilibrium condensation of a silicate gas in a H₂O–CO₂-rich vapour. We propose that the observed astrophysical carbonates have condensed in H₂O(g)–CO₂(g)-rich, high-temperature and high-density regions such as evolved stellar winds, or those induced by grain sputtering upon shocks in protostellar outflows.

Experiments were performed in a reaction chamber¹¹ where a gas with a controlled composition was produced by laser ablation of a glass target, with either a 'solar' composition¹² of calcium, aluminium, magnesium and silicon (MgO, SiO₂, Al₂O₃ and CaO: 36.1, 56.9, 4, 3 wt% respectively), or a Ca–Al-rich composition (MgO, SiO₂, Al₂O₃, CaO: 9.3, 48.6, 18.7, 23.4 wt%). Condensation of these silicate gases occurred in ambient gases (Ar, CO₂, or a mixture of CO₂ and H₂O) with total pressures varying from ~1 to 30 mbar, temperatures between 25 and ~130 °C, and for duration from ~40 to 464 min (Table 1). The condensed material, collected on plates located above the target was analysed by transmission electron microscopy (TEM) and mid- and far-infrared spectroscopy. Detailed descriptions of both the experimental device and the analytical procedure are given in the Supplementary Information. During our experiments, the hot laser-ablated gas expands in the cold ambient gas and decelerates rapidly owing to the relatively high pressure in the reaction chamber. The resulting supersaturation that takes place at the end of the shockwave yields non-equilibrium homogeneous condensation of ~10-nm-sized particles above the target^{13,14} (Fig. 1). TEM observations showed that the condensed nanoparticles are amorphous in all explored conditions (Fig. 1) except for the two experiments performed at 50 and 130 °C where nanoparticles are very slightly crystallized. In addition, material condensed at 25 °C either in 'dry' CO₂ (~4 mbar) or in 'wet' CO₂ (H₂O + CO₂, ~25 mbar) are not chemically fractionated (within analytical errors, ±20%) relative to the gas composition.

The mid-infrared spectra of products condensed in 'dry' CO₂ or Ar gas from Ca–Al-rich or 'solar' gas show that only amorphous silicates are present (for example, sample 119; Fig. 2). In contrast, spectra of materials condensed in 'wet' CO₂ gas show, in addition to the Si–O–Si stretching vibrations feature¹⁵, three features indicative of the presence of water and carbonate ions (for example, sample 130; Fig. 2). The two features corresponding to H₂O bending and stretching vibrations¹⁶ show the presence of highly bonded and trapped free water in the condensed material. The splitting of the carbonate stretching ν_3 feature¹⁷ into two main components (~1,430 and ~1,360 cm^{–1}) and the presence of the forbidden symmetric stretching ν_1 indicate that the carbonate ion does not exhibit strictly D_{3h} symmetry. These characteristics and the presence of highly bonded water thus show that the condensed carbonate is a complex hydrated carbonate with site distortions and several crystallographically distinct CO₃^{2–} groups in the structure. The far-infrared spectrum of Ca–Al-rich material condensed in 'wet' CO₂ (25 °C, 20 mbar H₂O + 4 mbar CO₂, for 493 min) shows the bands characteristics of the calcite lattice vibrations³ (Supplementary Information and Fig. 3). Mid- and far-infrared spectra of Ca–Al-rich condensate thus indicate that the condensation of a Ca–Al-rich silicate gas in 'wet' CO₂ yields formation of calcium-rich carbonates. Their subsequent annealing indicates also the crystallization of Ca-rich carbonates from the amorphous ones (Fig. 1). The similarity of the mid-infrared profiles of 'solar' and Ca–Al-rich condensates suggests that the carbonates are calcium-rich for both compositions (Supplementary Information). Electron energy loss spectroscopy showed that the repartition of these carbonates within the amorphous silicate matrix is not homogeneous. Using integrated absorption coefficients¹⁸ of 100–300 cm² g^{–1} for ν_3 and 1,500–3,000 cm² g^{–1} for the Si–O–Si stretching vibrations band, the relative amount of carbonate to silicate is 1 to 10% for both condensates.

The production of carbonate anions by collisions of the laser-ablated gas with CO₂(g) is excluded because water vapour is required for carbonate formation (for example, samples 130 and 119; Fig. 2). This indicates that the laser has no direct influence on carbonate formation. Carbonates do not form outside the apparatus by alteration of the nanoparticles in air or in the reaction chamber by surface reaction between liquid water, deposited as molecular layers of H₂O, and the nanoparticles. Indeed, several months of air exposure for the samples condensed in 'dry' CO₂ does not yield any formation of carbonates, whereas they form in an hour during silicate gas condensation with H₂O and CO₂ in atmospheric proportions (for example, sample 125; Table 1). Formation of carbonates by trapping and subsequent reaction of CO₂(g) and H₂O(g) in the silicate

¹Centre de Recherches Pétrographiques et Géochimiques—CNRS UPR 2300, 15 rue Notre Dame des Pauvres, BP 20, ²Ecole Nationale Supérieure de Géologie, INPL, BP40,

³Laboratoire Environnement et Minéralurgie—CNRS UMR 7569, 15 avenue du Charmois, BP40, 54501 Vandœuvre-les-Nancy, France. ⁴Laboratoire d'Etude de la Matière Extraterrestre—CNRS UMS 2679, Muséum National d'Histoire Naturelle, 61 rue Buffon, 75005 Paris, France. ⁵Institut d'Astrophysique Spatiale—CNRS UMR 8617, Université Paris-Sud, Bât. 121, 91405 Orsay, France. ⁶Service Commun de Microscopie Electronique à Transmission, Faculté des Sciences—Université Henri Poincaré, Nancy I, BP 239, 54506 Vandœuvre-les-Nancy, France. [†]Present address: Institute for Geophysics and Planetary Physics, Lawrence Livermore National Laboratory, 7000 East Avenue, L-413, Livermore, California 94550, USA.

Table 1 | Experimental conditions and mid-infrared characteristics of the samples condensed from Ca-Al-rich gas

Experiment	$p_{\text{H}_2\text{O}}$ (mbar)	Gas	p_{gas} (mbar)	Temperature (°C)	Duration (min)	Silicates (cm^{-1})	Ca/Si†	1,320/1,430‡	1,360/1,430‡	1,480/1,430‡	OH ₁ /Si§	OH ₂ /Si§	1,590/1,630‡
Glass target	-	-	-	-	-	970	0	-	-	-	-	-	-
Effect of water partial pressure													
119	0	CO ₂	4.6	25	42	980	0	-	-	-	-	-	-
143	10 ⁻³	CO ₂	1.2	25	54	1,020	0.10	0	0	0.58	0.02	0.74	-
172	1	CO ₂	0.3	35	56	1,012	0.09	0	0	0.10	0.04	1.09	0.96
130	20	CO ₂	1.8	25	55	1,022	0.44	0.28	0.52	0.13	0.20	2.29	0.48
Effect of condensation temperature													
204	20	CO ₂	3.5	25	464	1,031	0.45	0.19	0.10	0.28	0.13	3.11	0.04
124	20	CO ₂	4.6	50*	368	1,010	0.07	0	0.39	0	0.02	0.43	0.71
126	20	CO ₂	4.6	130*	240	1,000	0.18	0.08	0	0	0	0.48	-
Effect of CO ₂ partial pressure													
130	20	CO ₂	1.8	25	55	1,022	0.44	0.28	0.52	0.13	0.20	2.29	0.48
125	20	CO ₂	0.3	25	41	1,028	0.22	0.10	0.05	0.75	0.16	2.15	0

* Temperature ($\pm 30\%$) estimated from thermal gradient.
† Band area ratio with Ca and Si being the carbonate peak and the silicate peak.
‡ Band area ratio of carbonate deconvoluted bands (Fig. 2, Supplementary Information).
§ Band area ratio with OH₁ and OH₂ being to the 1,600 and 3,300 cm^{-1} water peaks, respectively (see Fig. 2, Supplementary Information).
The temperature and the duration of the condensation experiment are given. $p_{\text{H}_2\text{O}}$ and p_{gas} correspond to the partial pressure of water and of the additional ambient gas during the condensation experiment. The relative intensities of the bands show that the carbonate characteristics depend on the condensation conditions (see also Fig. 2 and Supplementary Information) as follows. First, a decrease in the partial pressure of water ($p_{\text{H}_2\text{O}}$) or an increase of the condensation temperature is associated with a decrease of total water and carbonate content, a decrease in free trapped water, and a crystallo-chemical change of the carbonates. Second, a decrease in the partial pressure of CO₂ yields a slight decrease in the amount of carbonates and of associated highly bonded water but also to a structural change of the carbonate.

nanostructure during condensation can also be discarded because of the absence of CO₂ trapping in ‘dry’ CO₂ experiments. Infrared studies showed that the carbonate characteristics depend on the condensation conditions (Table 1, Fig. 2 and Supplementary Information). Thus, carbonates form directly during the non-equilibrium condensation of the silicate nanoparticles, by a kinetically controlled chemical reaction involving CO₂(g), H₂O(g) and the silicate gas, with $p_{\text{H}_2\text{O}}$ being the limiting parameter. We propose that the reaction proceeds in two steps: (1) hydration in the gas phase of molecular clusters of Ca- or Mg-rich silicate and (2) reaction of these complexes with CO₂(g). During this reaction, Ca-rich carbonates are preferentially formed relative to Mg-rich carbonates. The change of the condensed material characteristics as a function of condensation temperature suggests that this reaction occurs at the ambient gas temperature because of the abrupt decrease of pressure and temperature at the end of the shockwave¹⁹.
Ca-rich carbonates have been detected^{3–5} in planetary nebulae

NGC 6302, NGC 6537 and in several low- and intermediate-mass protostars. Although not experimentally supported, several processes have been proposed^{3–5} to explain their presence: (1) interaction of the silicate grains with CO₂-H₂O-rich vapour, (2) interaction of the silicate grains with their CO₂-rich water-ice mantles, with an increase of the H₂O mobility by X-ray heating, and (3) direct condensation at high-temperature from a CaO-CO₂-rich gas. However, these processes are not likely to yield formation of the observed carbonates because the short lifetime of these environments^{3–5} and the low partial pressures of H₂O and CO₂ will not favour the interaction of the silicate with the ice or gaseous water, and because the presence of a gas highly enriched in CaO is unlikely. Here, based on our experimental results, we propose that the carbonates form by chemical reaction between a ‘hot’ silicate gas and a 300–500-K H₂O-CO₂-rich gas during a condensation process far from thermal equilibrium, in agreement with the observation of carbonates in stellar environments with pronounced disequilibrium conditions,

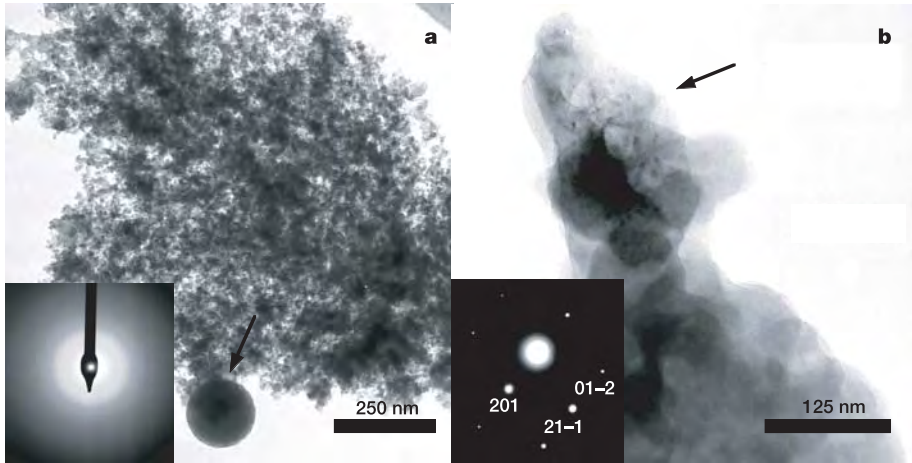


Figure 1 | Transmission electron micrographs of low-temperature condensates. TEM work was performed on a CM20 at the University Henri Poincaré, Nancy, France. **a**, Bright-field micrograph of the fluffy condensates (run product from experiment 204) condensed at 25 °C in ~20 mbar H₂O and 3.5 mbar CO₂ over 464 min. Inset, the diffraction pattern shows the amorphous structure of the condensates. The arrow indicates one ablation droplet ejected from the heated glass target subsurface during ablation. **b**, Bright-field micrograph of condensates condensed at 25 °C in ~20 mbar

H₂O and 1.8 mbar CO₂ (run product from experiment 130) and annealed in 0,001 bar of Ar at 65 °C over 10 days. The annealing of the condensed material yields compaction of the nanoparticles and crystallization of nanodomains. The arrow shows a crystallized part of the condensate. Inset, the associated diffraction pattern, which is close to that of calcite, and the chemical analysis (CaO-rich, 75% O and 25% Ca) of these nanodomains suggest the presence of crystalline Ca-rich carbonate.

such as protostars and descendants of the fairly massive stars¹⁸, NGC 6302 and NGC 6537. In evolved stellar or protostellar environments, 'hot' silicate gases are present, as well as water vapour^{20,21} and very likely CO₂ (refs 22–24). Although pressures in these environments are lower than in our experiments, timescales are longer and the occurrence of shockwaves or turbulences could yield high-density regions with supersaturation levels required to

produce non-equilibrium condensation. Note that iron, expected in circumstellar silicate gases, should not enter significantly in the carbonate structure, similarly to magnesium.

In evolved stellar winds, carbonates can form at around 300–500 K by non-equilibrium condensation of the hot silicate gas, which has not fully condensed at higher temperature owing to the rapid expansion of the gas⁸. Such a process could be favoured during the post-Asymptotic Giant Branch phase, when high-velocity winds interact with the high-density clumpy remnant envelope shaped by the previous slow winds. In descendants of massive AGB stars, the rapid stellar evolution results in cold dense circumstellar material close to the hot central star favouring non-equilibrium conditions⁸. Although calculations of condensation⁸ based on chemical equilibrium indicate that the amount of carbonates formed in evolved stars should be very low, our mechanism involving a non-equilibrium reaction results in a higher rate of carbonate formation. In protostellar outflows, highly energetic radiation, grain–grain or ion/atom–grain collisions associated with shocks lead²⁵ to the sputtering of silicate grains as shown by the observations of large gaseous SiO enhancements (for example, a column density of $\sim 10^{11}$ – 10^{13} cm⁻²). In addition, numerous warm and dense regions enriched in water vapour (~ 300 K) have been observed²⁶ in the inner protostar envelopes. Because our experiment mimics the sputtering of grains by stellar winds and the expansion of the produced hot silicate gas in a H₂O–CO₂-rich region, we propose that carbonates condense in the regions where the outflows interact with the innermost part of the protostellar envelope²⁷. Finally, the strong emission

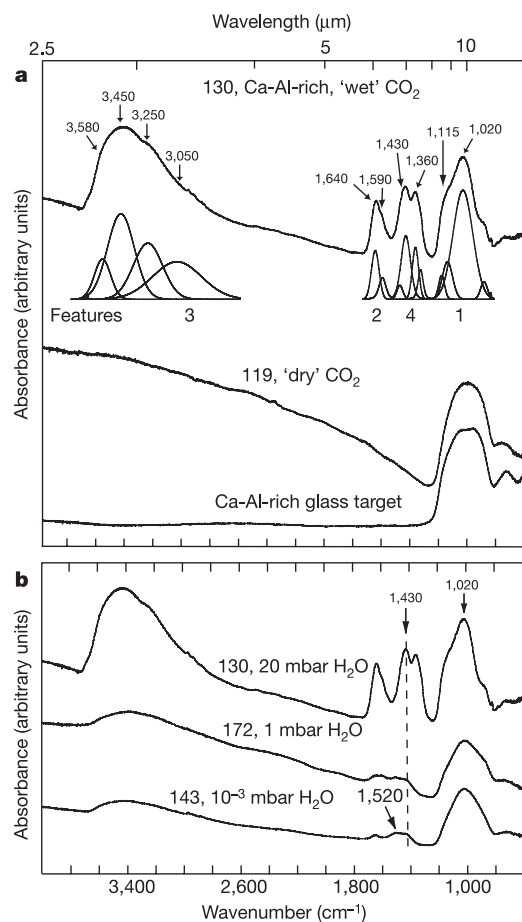


Figure 2 | Typical mid-infrared transmission spectra (4,000–600 cm⁻¹) of material condensed from Ca-Al-rich gas. **a**, Infrared-spectra from Ca-Al-rich glass target, material condensed in 'dry' CO₂ gas at 25 °C (sample 119) and material condensed at 25 °C in 'wet' CO₂ from Ca-Al-rich gas (sample 130). The spectra of sample 119 and the glass target are similar and show only the Si–O–Si stretching vibrations feature at ~ 980 cm⁻¹. Sample 130 shows a spectrum with features arising from the stretching mode of the silicates (feature 1, $\sim 1,020$ cm⁻¹), the stretching (feature 3, $\sim 3,300$ cm⁻¹) and bending (feature 2, $\sim 1,600$ cm⁻¹) mode of water and the carbonate doubly degenerate asymmetric stretching ν_3 (feature 4, $\sim 1,400$ cm⁻¹). Note also the presence of the carbonate forbidden symmetric stretching ν_1 ($\sim 1,115$ cm⁻¹). Peaks resulting from the fitting of the silicate, ν_3 and water bands are shown on the spectrum of sample 130. Feature 4 was best-fitted with four peaks at $\sim 1,470$, $\sim 1,430$, $\sim 1,360$ and $\sim 1,320$ cm⁻¹. Fits of features 2 and 3 show the presence of highly bonded water ($\sim 1,590$, $3,250$ and $3,050$ cm⁻¹) and trapped free water ($\sim 1,640$, $3,580$ and $3,450$ cm⁻¹). **b**, Infrared-spectra of material condensed at 25 °C for ~ 50 min under different water partial pressure ($p_{\text{H}_2\text{O}}$) (Samples 130, 172 and 143; Table 1). A $p_{\text{H}_2\text{O}}$ decrease is associated with a decrease of water and carbonate content and a crystallo-chemical change of the carbonates with a decrease of the $1,360$ cm⁻¹ band and the apparition of a band around $1,520$ cm⁻¹. Note that carbonates form even if the partial pressure of water is lower than the equilibrium vapour pressure. The shift of the Si–O–Si feature from 980 to $1,020$ cm⁻¹ in the different spectra is related to the polymerization degree of the silicate tetrahedrons (for example, amount of alkaline-earth elements, presence of water) and to the presence of the ν_1 band.

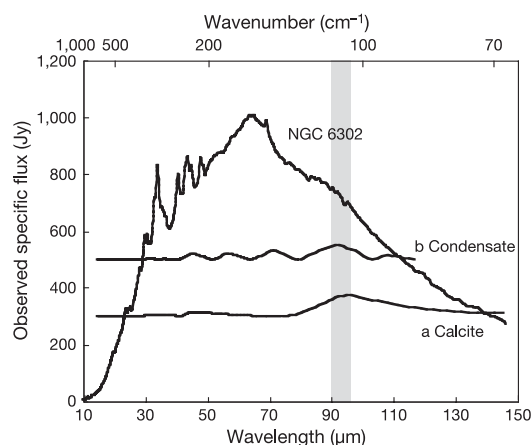


Figure 3 | Spectrum of the NGC 6302 planetary nebulae as observed by the Infrared Space Observatory Short Wavelength and Long Wavelength Spectrometers (10–200 μm). The full spectrum can be reproduced by the sum of warm C-rich grains (118 K) and cold O-rich grains comprising forsterite, diopside, water ice and enstatite in addition to amorphous silicate (94%). The unidentified line at $91 \mu\text{m}$ has been interpreted as resulting from the presence of calcite which shows a band at $92 \mu\text{m}$ (spectrum a). Spectrum b gives the infrared signature of material condensed at 25 °C in ~ 20 mbar H₂O and ~ 4 mbar CO₂ over 493 min. These two far-infrared (700 – 50 cm⁻¹) spectra (a and b) have been measured at Institut d'Astrophysique Spatiale, Orsay, using a Bruker 66V Fourier Transform InfraRed spectrometer (resolution 4 cm⁻¹). Measurements were conducted in transmission mode on material embedded in a polyethylene window (dilution 6–8%). Their expected emission at 50 K (spectra a and b) has been obtained by multiplying the baseline corrected spectra by a Planck function at 50 K. The intensity scales for both calcite and condensed material are arbitrary. The $92\text{-}\mu\text{m}$ band of planetary nebulae spectrum is well fitted by the carbonate band of the amorphous condensate. The additional presence of the weak 44 -, 56 - and $71\text{-}\mu\text{m}$ bands in the condensed material spectra has no large influence on the fitting with the planetary nebulae spectra since these bands mainly overlapped with the large planetary nebulae 40 - and $60\text{-}\mu\text{m}$ bands whose identification—although certainly related to the presence of enstatite, crystalline H₂O, diopside and unidentified species—is still not clear²⁹.

of X-rays by protostars may be associated²⁸ with energetic processes such as shocks or highly energetic winds inducing the sputtering of the silicate grains in the inner envelope and the production of a hot silicate gas, which may in turn condense in the H₂O-rich region.

Contrary to previous assumptions^{3–5}, this new process using only gas-phase species can thus produce carbonates in both types of environments with carbonate-to-silicate ratios close to those observed^{3–5}. It explains the presence of amorphous silicates²⁹ and that of carbonates, which could be amorphous rather than crystalline (Fig. 3). We propose that this mechanism of formation by condensation out of equilibrium is widespread in energetic H₂O(g)-rich circumstellar environments. In such a case, in addition to the present detection of the ~95- μ m emission feature, the main vibrational mode of carbonate (~7 μ m) must be detected in absorption towards these objects, using the much greater sensitivity of the Spitzer infrared satellite. Such a detection will not imply the presence of liquid water. Finally, our experiments establish that non-equilibrium vapour phase condensation is an important mechanism of dust formation in astrophysical environments and could result in the formation of species that are not expected by equilibrium condensation³⁰.

Received 13 January; accepted 3 August 2005.

- Sandford, S. A. Acid dissolution experiments: carbonates and the 6,8- μ m bands in interplanetary dust particles. *Science* **231**, 1540–1541 (1986).
- Lewis, J. S. Low-temperature condensation from the solar nebula. *Icarus* **16**, 241–252 (1972).
- Kemper, F. *et al.* Detection of carbonates in dust shells around evolved stars. *Nature* **415**, 295–297 (2002).
- Ceccarelli, C. *et al.* Discovery of calcite in the solar type protostar NGC 1333-IRAS4. *Astron. Astrophys.* **395**, L29–L33 (2002).
- Chiavassa, A., Ceccarelli, C., Tielens, A. G. G. M., Caux, E. & Maret, S. The 90–110 μ m dust features in low to intermediate mass protostars: calcite? *Astron. Astrophys.* **432**, 547–557 (2005).
- Onaka, T. & Okada, Y. Detection of far-infrared features in star-forming regions. *Astrophys. J.* **585**, 872–877 (2003).
- Hofmeister, A. M., Wopenka, B. & Locock, A. J. Spectroscopy and structure of hibonite, grossite, and CaAl₂O₄: Implications for astronomical environments. *Geochim. Cosmochim. Acta* **68**, 4485–4503 (2004).
- Ferrarotti, A. S. & Gail, H.-P. Mineral formation in stellar wind V. Formation of calcium carbonate. *Astron. Astrophys.* **430**, 959–965 (2005).
- Tielens, A. G. G. M., Waters, L. B. F. M., Molster, F. J. & Justtanont, K. Circumstellar silicate mineralogy. *Astrophys. Space Sci.* **255**, 415–426 (1998).
- Gail, H.-P. & Sedlmayr, E. Mineral formation in stellar winds. I. Condensation sequence of silicate and iron grains in stationary oxygen rich outflows. *Astron. Astrophys.* **347**, 594–616 (1999).
- Toppani, A., Libourel, G., Robert, F., Ghanbaja, J. & Zimmermann, L. Synthesis of refractory minerals by high-temperature condensation of a gas of solar composition. *Lunar Planet. Sci. Conf.* XXXV, abstr. 1726 (2004).
- Anders, E. & Grevesse, N. Abundances of the elements: meteoritic and solar. *Geochim. Cosmochim. Acta* **53**, 197–214 (1989).
- Abraham, F. F. *Homogeneous Nucleation Theory* 263 (Academic, New York, 1974).
- Chen, L.-C. in *Pulsed Laser Deposition of Thin Films* (eds Chrisey, D. B. & Hubler, G. K.) 167–198 (John Wiley & Sons, New York, 1994).
- McMillan, P. F. & Wolf, G. H. in *Structure, Dynamics and Properties of Silicate Melts* (eds Stebbins, J. F., McMillan, P. F. & Dingwell, D. B.) 247–315 (Mineralogical Society of America, Washington, USA, 1995).
- Ihinger, P. D., Hervig, R. L. & McMillan, P. F. in *Volatiles in Magmas* (eds Carroll, M. R. & Holloway, J. R.) 67–121 (Mineralogical Society of America, Washington, 1994).
- White, W. B. in *The Infrared Spectra of Minerals* (ed. Farmer, V. C.) 227–284 (Mineralogical Society, London, 1974).
- Kemper, F., Molster, F. J., Jäger, C. & Waters, L. B. F. M. The mineral composition and spatial distribution of the dust ejecta of NGC 6302. *Astron. Astrophys.* **394**, 679–690 (2002).
- Wood, R. F., Leboeuf, J. N., Chen, K. R., Geohegan, D. B. & Poretzki, A. A. Dynamics of plume propagation, splitting, and nanoparticle formation during pulse-laser ablation. *Appl. Surf. Sci.* **127–129**, 151–158 (1998).
- Payne, H. E., Philips, J. A. & Terzian, Y. A young planetary nebula with OH molecules: NGC 6302. *Astrophys. J.* **326**, 368–375 (1988).
- Marret, S., Ceccarelli, C., Caux, E., Tielens, A. G. G. M. & Castets, A. Water emission in NGC 1333-IRAS4. The physical structure of the envelope. *Astron. Astrophys.* **395**, 573–585 (2002).
- Cami, J. *et al.* CO₂ emission in EP Aqr: Probing the extended atmosphere. *Astron. Astrophys.* **360**, 562–574 (2000).
- d'Hendecourt, L. B. & Jourdain de Muizon, M. The discovery of interstellar carbon dioxide. *Astron. Astrophys.* **223**, L5–L8 (1989).
- de Graauw, T. *et al.* SWS observations of solid CO₂ in molecular clouds. *Astron. Astrophys.* **315**, L345–L348 (1996).
- Nisini, B., Codella, C., Giannini, T. & Richer, J. S. Observations of high-J SiO emission along the HH211 outflow. *Astron. Astrophys.* **395**, L25–L28 (2002).
- van Dishoeck, E. F. ISO spectroscopy of gas and dust: from molecular clouds to protoplanetary disks. *Annu. Rev. Astron. Astrophys.* **42**, 119–167 (2004).
- Gueth, F., Guilloteau, S., Dutrey, A. & Bachiller, R. Structure and kinematics of a protostar: mm-interferometry of L 1157. *Astron. Astrophys.* **323**, 943–952 (1997).
- Feigelson, E. D. & Montmerle, T. High-energy processes in young stellar objects. *Annu. Rev. Astron. Astrophys.* **37**, 363–408 (1999).
- Molster, F. J. *et al.* The complete ISO spectrum of NGC 6302. *Astron. Astrophys.* **372**, 165–172 (2001).
- Grossman, L. Condensation in the primitive nebula. *Geochim. Cosmochim. Acta* **36**, 597–619 (1972).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements L. Zimmerman, R. Ruppen and F. Bendisari are thanked for technical help and J. Aléon, E. Deloule and J. Bradley for discussions. C. Kemper is thanked for providing the spectral data of NGC 6302.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.T. (toppani2@lnl.gov).

Secondary craters on Europa and implications for cratered surfaces

Edward B. Bierhaus¹, Clark R. Chapman² & William J. Merline²

For several decades, most planetary researchers have regarded the impact crater populations on solid-surfaced planets and smaller bodies as predominantly reflecting the direct ('primary') impacts of asteroids and comets¹. Estimates of the relative and absolute ages of geological units on these objects have been based on this assumption². Here we present an analysis of the comparatively sparse crater population on Jupiter's icy moon Europa and suggest that this assumption is incorrect for small craters. We find that 'secondaries' (craters formed by material ejected from large primary impact craters) comprise about 95 per cent of the small craters (diameters less than 1 km) on Europa. We therefore conclude that large primary impacts into a solid surface (for example, ice or rock) produce far more secondaries than previously believed, implying that the small crater populations on the Moon, Mars and other large bodies must be dominated by secondaries. Moreover, our results indicate that there have been few small comets (less than 100 m diameter) passing through the jovian system in recent times, consistent with dynamical simulations^{3–6}.

Surface ages can be derived from the spatial density of craters, but this association presumes that the craters are made by interplanetary impactors, arriving randomly in time and location across the surface. Secondary craters cause confusion because they contaminate the primary cratering record by emplacing large numbers of craters, episodically, in random and non-random locations on the surface. The number and spatial extent of secondary craters generated by a primary impact has been a significant research issue^{7–13}. If many or most small craters on a surface are secondaries, but are mistakenly identified as primaries, derived surface ages or characteristics of the impacting population size–frequency distribution (SFD) will be in error. Accurately identifying secondary craters and removing their contribution from the primary crater SFD would significantly improve confidence in calculations of surface ages and impactor fluxes, which are the only chronometers for planetary surfaces for which we do not have physical samples (that is, most Solar System objects). Indeed, our understanding of surface ages on unsampled regions of the Moon and for Mars is based almost entirely on the 'crater chronology' (the correlation between relative surface age and observed crater density). Thus determining the contribution of secondaries is critical.

Identifying secondaries adjacent to their parent primary has been done^{14,15} on the basis of unusual morphologies and distinct spatial patterns of proximate secondaries. But the extent of distant, far-field secondaries (those far from their parent primary) has previously been indeterminate because most planetary surfaces are too densely cratered to disentangle the small crater population into primary and secondary. The first high-resolution Galileo spacecraft images of Europa, however, revealed a crater population with densities much lower than those for the Moon or most regions of Mars. Also, these

craters were typically not spatially random, but instead appeared in clumps and clusters¹⁶ even at distances far (several hundred kilometres) from the nearest large primary crater. This clustered spatial distribution contrasts with primary impacts, which are spatially random. The low spatial density and unusual non-random spatial distribution of Europa's small craters enabled us to achieve what has been previously difficult, namely unambiguous identification and quantification of the contribution of distant secondary craters to the total crater SFD. This, in turn, allows us to re-examine the overall crater age-dating methodology. We discuss the specifics of the Europa data first.

We measured more than 17,000 craters in 87 low-compression, low-sun, high-resolution Europa images (scales $<60 \text{ m pixel}^{-1}$), which cover nine regions totalling 0.2% of Europa's surface (a much larger percentage of Europa has been imaged at lower resolutions), plus 10,000 more secondaries surrounding the large, well-imaged craters Tyre ($\sim 44 \text{ km}$ diameter) and Pwyll ($\sim 25 \text{ km}$ diameter). Most of the SFDs for the small, far-field craters have 'steep slopes', which means that they have a differential power-law size index of $b < -4$. (The differential SFD follows the form $dN = kD^b dD$, where dN is the number of craters of diameter D in diameter range dD , and k and b are constants. We use a χ^2 nonlinear fit to determine k and b .)

We find that the small craters are spatially clustered. To estimate the random (primary) background population, we developed a novel algorithm¹⁷ that removes the strongly clustered (secondary) craters, simultaneously identifying specific clusters and any spatially random population. The method combines the single-linkage¹⁸ (SLINK) hierarchical clustering algorithm, Monte Carlo methods, and a clustering parameter. We divide the crater population into groups on the basis of their probabilities P of non-randomness, where a cluster with $P > 2\sigma$ indicates that the cluster is non-random with significance greater than two standard deviations (2σ , or 95%), and thus has only a 5% chance of occurring by random impacts. The algorithm first generates a suite of random populations (mimicking primary cratering) of craters that possess the same spatial density as the data. It then applies SLINK to each population, calculating a clustering parameter during each step of the clustering process. The clustering parameter (κ) evaluates the 'clusteredness' of a point on the basis of its local spatial density (that is, an object in a cluster has many close neighbours, not just one or two), the global density of the mosaic, and the similarity between the local and global densities. The algorithm then compares the κ values of the Europa data with those of the simulations of random impacts. Among nine regions that we studied, the minimum, median and maximum percentage of craters strongly clustered at probabilities of non-randomness $P > 2\sigma$ is 35%, 71% and 80%, respectively. Figure 1 exemplifies our results.

We find that most of the moderately-clustered (at probabilities of non-randomness of $1\sigma < P \leq 2\sigma$, see Fig. 1) and spatially random craters are also secondaries; the actual percentage of true primaries

¹Lockheed Martin, Space Exploration Systems, MS 58110, PO Box 179, Denver, Colorado 80201, USA. ²Southwest Research Institute, 1050 Walnut Street, Suite 400, Boulder, Colorado 80302, USA.

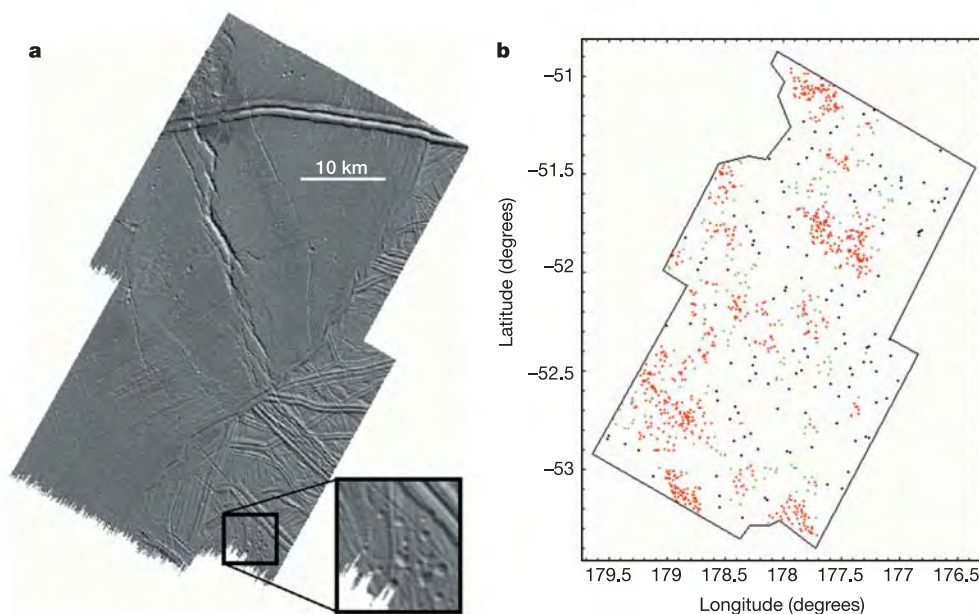


Figure 1 | A sample region on Europa that demonstrates the extreme 'clusteredness' of Europa's small craters. a, The $\sim 43 \text{ m pixel}^{-1}$ E17LIBLIN01 Galileo mosaic of Europa. North is up, and the Sun is to the upper-right. About 1,000 craters are identified. The inset zooms in on a portion of one cluster. **b,** Plot of crater-centre locations, colour-coded to

show the degree of clustering: red points are spatially non-random with probability $P > 2\sigma$ (strongly clustered), green points are spatially non-random with $1\sigma < P \leq 2\sigma$ (moderately clustered), and black points are spatially random. The corresponding percentages are 71%, 15% and 14%, respectively.

may be only a few per cent or less. There are several reasons for this conclusion. First, the craters clustered at lower significances are, in aggregate, statistically still clustered. Second, the SFDs of the spatially random craters do not exhibit a uniform shape (that is, a similar exponent b), as would be expected for craters caused by interplanetary impactors. Instead, they vary from region to region, which is not consistent with a single impacting flux. For example, one region's spatially random population has $b = -5.4 \pm 1.0$, while another region's spatially random population has $b = -2.2 \pm 0.7$. Third, evidence for unclustered secondaries comes from some crater populations that are spatially coincident with crater rays. The best example is a region roughly 1,000 km distant from the fresh impact crater Pwyll that contains a bright Pwyll ray (a known ejecta product from primary cratering). The spatial coincidence¹⁹ between the small craters in this region and the Pwyll ray assures us that these are secondaries, yet within the ray the strongly clustered craters are 'only' 60% of the population. Fragments that make distant secondary craters have longer flight times (in this case, tens of minutes), which give the fragments more time to spatially separate.

Therefore, the percentage of spatially random craters in our study regions represent a strong upper limit on the percentage of small craters that are primaries; among the nine regions studied, the median percentage of spatially random craters is 10%. But we can refine this estimate because a certain fraction of the spatially-random craters are in fact secondaries. Our conservative evaluation is that at least 50% of the spatially random craters need to be secondaries to account for the observed variation in SFDs and spatial density (from 0.01 to 0.17 craters km^{-2}) between measured regions—that is, at least half must be secondaries to account for a variation in b from -2.2 to -5.4 . We note that all regions (visible in the available high-resolution data) contain multiple crater clusters, and that the spatial density of the random population is directly correlated with the spatial density of the clustered population.

There are (at least) three significant implications of our discovery that the bulk of Europa's small crater population consists of secondary craters. (1) Using the more numerous, and thus statistically more significant, small craters to derive surface ages of small geological

units is now suspect for Europa. Far-flung secondaries from a single, distant primary crater can dominate a local crater population, so a high-density crater population does not necessarily reflect a greater integrated exposure time; it could instead result from a single, possibly recent, exposure to a cluster of ejecta from a single impact, in an instant of time. (2) The jovian system's primary impacting population, which is dynamically understood to be mostly comets derived from the Kuiper belt/scattered disk^{3–6,20,21}, is evidently deficient in objects smaller than 100 m diameter (that is, the size distribution is shallower than expected from extrapolations of comet SFDs). Fewer small primary impacts indicate fewer cometary impactors. This may be the first observational constraint for the Kuiper belt at those diameters; perhaps small comets are rarely made, or perhaps some process depletes them before reaching the jovian system. (3) On Europa, the production of ejecta that make secondary craters is unexpectedly efficient. ('Secondary cratering efficiency' refers to the fraction of ejected material that generates secondaries for a given primary impact energy.) There are very few large primary craters on Europa (less than 40 with diameter $> 10 \text{ km}$ are currently known globally²² as seen in the lower resolution images), yet our identification of tens of thousands of secondaries in only $\sim 0.2\%$ of Europa's surface imaged at high resolution implies that the few large primaries made at least 10 million secondaries larger than 100 m. The steep size distributions suggest yet more numerous craters at smaller diameters.

At least for Europa, our work unequivocally resolves the relative abundance of primary and secondary craters. We now return to questions of crater distributions in the Solar System in general, and the implications for impactor size distributions and age dating. Laboratory-scale experiments demonstrate that impacts into ice and rock targets yield ejecta fragments with steep SFDs, with ice-impacts generating ejecta more efficiently^{23–25}. We note that the laboratory ice-impact experiments took place at temperatures warmer (255 K) than Europa's surface ($\sim 110 \text{ K}$), and at lower impact velocities (tens of m s^{-1} to 1 km s^{-1} , compared with the $\sim 20 \text{ km s}^{-1}$ average comet impact velocity³). The Moon and Mars both possess a steep branch in their crater SFD at sizes below a few kilometres^{26,27}.

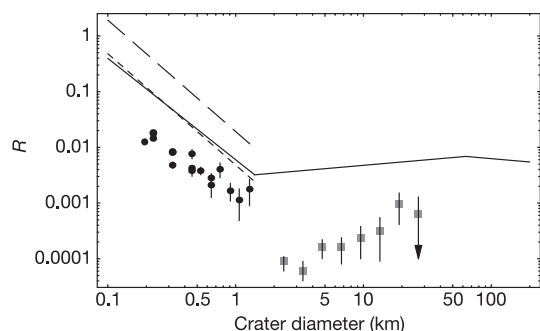


Figure 2 | A comparison of european and lunar crater size-frequency distributions, demonstrating that lunar secondaries could account for most of the observed small lunar craters. The size–frequency distributions (SFDs) are given in R format¹; R is the differential distribution divided by D^{-3} , where D is the crater diameter. The grey squares are the measured²² european primary crater population for diameters greater than 2 km (diameters least affected by secondaries). The filled circles represent some of our measurements from high-resolution images of Europa, including data from the crater population that appears in Fig. 1 (the error bars for both populations are the standard $\sqrt{N}$, representing one standard deviation for counting statistics). The solid black line is the average lunar maria crater SFD (taken from ref. 26). The long-dashed line is the calculated secondary crater population generated by the lunar post-mare primary crater population with diameters $10 \text{ km} < D < 64 \text{ km}$, assuming the same secondary cratering efficiency we measured on Europa. This is clearly too high, generating a small crater population that not only meets but exceeds the observed lunar population. The short-dashed line is the calculated secondary population with the european efficiency decreased by a factor of about four (consistent with laboratory-scale impact experiments²⁴).

We integrated our derived secondary cratering efficiency over the visible large ($10 \text{ km} < D < 64 \text{ km}$), post-mare, primary impact craters on the Moon. We calculate that the steep branch of the lunar crater SFD at small diameters could be fully accounted for by secondary cratering, even if the Moon's secondary cratering efficiency is less than Europa's—that is, even if rock targets generate ejecta less efficiently than ice targets (Fig. 2).

Our work raises doubts regarding methods that use the lunar small-crater distribution to calibrate other inner Solar System surface ages (for example, Mars^{28,29}). If, as on Europa, lunar and martian secondaries are 95% of the small crater (less than a few kilometres) population, the error bars (and thus derived surface ages) on any residual primary crater population become large (uncertainties are 20 times the measured density value). This uncertainty applies to both the measured population on a martian surface unit and the lunar SFD that supposedly represents absolute age. We emphasize that traditional age-dating analyses still derive robust ages when using large craters (greater than a few kilometres diameter), which are less likely to be secondaries. However, the technique becomes increasingly unreliable when applied to dating tiny geographical units using small craters, which may be mostly secondaries.

Our discovery that even spatially random crater populations contain numerous secondaries confounds the general rule-of-thumb that one can 'avoid' secondaries by excluding obviously clustered craters from measurements and analysis. Rather, secondaries are intimately mixed, both non-randomly and randomly, into crater distributions on planetary surfaces. The ability of a few large impacts to distribute a globally extensive secondary crater population demonstrates that one cannot use distance from a primary to minimize contributions from secondaries.

Finally, any attempt to derive surface ages or traits of impacting populations (that is, of near-Earth objects at diameters below what is robustly sampled in telescopic surveys) based on lunar or martian small-crater statistics may suffer a significant and perhaps uncorrectable bias due to the contribution from secondaries. Indeed, other

research³⁰ finds that a single 10 km primary contributed $\sim 10^7$ secondary craters from 10 m to 100 m diameter to the martian crater distribution.

Received 1 April; accepted 20 July 2005.

- Basaltic Volcanism Study Project. In *Basaltic Volcanism on the Terrestrial Planets* (eds Kaula, W. M. et al.) Ch. 8 (Pergamon, New York, 1981).
- Shoemaker, E. M., Hackman, R. J. & Eggleton, R. E. Interplanetary correlation of geologic time. *Adv. Astronaut. Sci.* **8**, 70–89 (1963).
- Zahnle, K., Schenk, P. & Levison, H. Cratering rates in the outer Solar System. *Icarus* **163**, 263–289 (2003).
- Zahnle, K., Dones, L. & Levison, H. Cratering rates on the Galilean satellites. *Icarus* **136**, 202–222 (1998).
- Levison, H. F. et al. Planetary impact rates from ecliptic comets. *Icarus* **143**, 415–420 (2000).
- Levison, H. F. & Duncan, M. J. From the Kuiper Belt to Jupiter-family comets: the spatial distribution of ecliptic comets. *Icarus* **127**, 13–32 (1997).
- Shoemaker, E. M. Ballistics of the Copernican ray system. *Proc. Lunar Planet. Colloq.* **2**, 7–21 (1960).
- Neukum, G., Ivanov, B. A. & Hartmann, H. K. Cratering records in the inner solar system in relation to the lunar reference system. *Space Sci. Rev.* **96**, 55–86 (2001).
- Fielder, G. Ray elements and secondary-impact craters on the Moon. *Astrophys. J.* **135**, 632–637 (1962).
- Roberts, W. A. Secondary craters. *Icarus* **3**, 348–364 (1964).
- Kopal, Z. The nature of secondary craters photographed by Ranger VII. *Icarus* **5**, 201–213 (1966).
- Lucchitta, B. K. Crater clusters and light mantle at the Apollo 17 site—A result of secondary impacts from Tycho. *Icarus* **30**, 80–96 (1977).
- Shoemaker, E. M. Preliminary analysis of the fine structure of the lunar surface. In *Ranger VII, Part II, Experimenters' Analyses and Interpretations* 75–134 (Tech. Rep. No. 32–700, Jet Propulsion Laboratory, Pasadena, 1965).
- Vickery, A. M. Variation in ejecta size with ejection velocity. *Geophys. Res. Lett.* **14**, 726–729 (1987).
- Vickery, A. M. Size-velocity distribution of large ejecta fragments. *Icarus* **67**, 224–236 (1986).
- Bierhaus, E. B., Chapman, C. R. & Merline, W. J. On the clustering of Europa's small craters. *Lunar Planet. Sci. Conf. XXXII*, abstr. no. 1967 (2001).
- Bierhaus, E. B. *Discovery that Secondary Craters Dominate Europa's Small Crater Population* PhD thesis, Univ. Colorado at Boulder (2004).
- Duda, R. O. & Hart, P. E. *Pattern Classification and Scene Analysis* (Wiley and Sons, New York, 1973).
- Bierhaus, E. B., Chapman, C. R., Merline, W. J., Brooks, S. M. & Asphaug, E. Pwyll secondaries and other small craters on Europa. *Icarus* **153**, 264–276 (2001).
- Duncan, M., Levison, H. & Dones, L. in *Comets II* (eds Festou, M. C., Keller, H. U. & Weaver, H. A.) 193–204 (Univ. Arizona Press, Tucson, 2004).
- Berstein, G. M. et al. The size distribution of trans-Neptunian bodies. *Astron. J.* **128**, 1364–1390 (2004).
- Schenk, P., Chapman, C. R., Zahnle, K. & Moore, J. M. in *Jupiter* (eds Bagenal, F., Dowling, T. E. & McKinnon, W. B.) 427–456 (Cambridge University Press, Cambridge, 2004).
- Kato, M. et al. Ice-on-ice impact experiments. *Icarus* **113**, 423–441 (1995).
- Arakawa, M. et al. Ejection velocity of ice fragments. *Icarus* **118**, 341–354 (1995).
- Kato, M. et al. Shock pressure attenuation in water ice at a pressure below 1 GPa. *J. Geophys. Res.* **106**, 17567–17578 (2001).
- Ivanov, B. A., Neukum, G., Bottke, W. F. & Hartmann, W. K. in *Asteroids III* (eds Bottke, W. F., Cellino, A., Paolicchi, P. & Binzel, R. P.) 89–101 (Univ. Arizona Press, Tucson, 2002).
- Hartmann, W. K. Martian cratering VI. Crater count isochrons and evidence for recent volcanism from Mars Global Surveyor. *Meteorit. Planet. Sci.* **34**, 167–177 (1999).
- Hauber, E. et al. Discovery of a flank caldera and very young glacial activity at Hecates Tholus, Mars. *Nature* **434**, 356–361 (2005).
- Murray, J. B. et al. Evidence from the Mars Express High Resolution Stereo Camera for a frozen sea close to Mars' equator. *Nature* **434**, 353–356 (2005).
- McEwen, A. S. et al. The rayed crater Zunil and interpretations of small impact craters on Mars. *Icarus* **176**, 351–381 (2005).

Acknowledgements E.B.B. was previously at the University of Colorado at Boulder and Southwest Research Institute, Boulder. NASA's Galileo and JSDAP programs funded the research reported here. We thank B. Ivanov for providing comments that improved the clarity and structure of this manuscript.

Author Information Reprints and permissions information is available at ngp.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to E.B.B. (edward.b.bierhaus@lmco.com).

LETTERS

Enhancing semiconductor device performance using ordered dopant arrays

Takahiro Shinada¹, Shintaro Okamoto², Takahiro Kobayashi² & Iwao Ohdomari^{1,2}

As the size of semiconductor devices continues to shrink, the normally random distribution of the individual dopant atoms within the semiconductor becomes a critical factor in determining device performance—homogeneity can no longer be assumed^{1–5}. Here we report the fabrication of semiconductor devices in which both the number and position of the dopant atoms are precisely controlled. To achieve this, we make use of a recently developed single-ion implantation technique^{6–9}, which enables us to implant dopant ions one-by-one into a fine semiconductor region until the desired number is reached. Electrical measurements of the resulting transistors reveal that device-to-device fluctuations in the threshold voltage (V_{th} ; the turn-on voltage of the device) are less for those structures with ordered dopant arrays than for those with conventional random doping. We also find that the devices with ordered dopant arrays exhibit a shift in V_{th} relative to the undoped semiconductor, that is twice that for a random dopant distribution (-0.4 V versus -0.2 V); we attribute this to the uniformity of electrostatic potential in the conducting channel region due to the ordered distribution of dopant atoms. Our results therefore serve to highlight the improvements in device performance that can be achieved through atomic-scale control of the doping process. Furthermore, ordered dopant arrays of this type may enhance the prospects for realizing silicon-based solid-state quantum computers¹⁰.

Doping of impurity atoms into semiconductors is essential to achieving the proper function of semiconductor devices through the control of electrical characteristics¹¹. So far, the semiconductor has been assumed to be homogeneously doped in the active channel region. In nanoscale semiconductor devices, however, the channel region will contain few dopant atoms and the assumption of uniform dopant distribution is no longer feasible. In this situation, the statistical fluctuation in dopant atom number due to a random Poisson distribution causes serious fluctuation in the device's functioning.

In order to suppress the conductance fluctuation due to the fluctuation in the number of dopant atoms, we have previously attempted to tailor the conductance of submicrometre resistors, which corresponds to the channel region in semiconductor devices, by one-by-one implantation of dopant ions, which we refer to as single ion implantation (SII). In SII, single ions are extracted by chopping a focused ion beam using a small aperture and high frequency beam deflection, and the number of implanted ions is controlled one-by-one by detecting secondary electrons emitted from a target outside upon a single ion incidence. A broad range of ion species such as Be, B, Si, P, Fe, Co, Ni, Cu, Ga, Ge, As, Pd, In, Sb, Pt and Au can now be implanted one-by-one with an aiming precision of 60 nm (ref. 9) and with potentially higher accuracy by remodelling the focused ion beam optics for SII. The number of single ions necessary to tailor the conductance value to a certain value on the

higher side of the initial distribution was implanted in each resistor. The initial conductance fluctuation (the ratio of the standard deviation to the average value of conductance from 22 resistors) of 63% was reduced to only 13% (ref. 8). By analysing the origin of the fluctuation after SII, the residual fluctuation turned out to be the fluctuation in the dopant atom position. Thus we have found that the control of not only the dopant atom number but also its position is essential.

We note that a solid-state quantum computer has been proposed by Kane¹⁰. Kane's original proposal requires single phosphorus atoms to be placed in an array in a Si layer beneath an insulating oxide layer. However, the incorporation of the dopant array is a major problem that has yet to be overcome. In this Letter, we demonstrate the fabrication of a semiconductor with an ordered array of dopant atoms.

Figure 1a sketches a simplified version of the resistor. The channel size is 300 nm wide, 3.2 μ m long and 90 nm thick employing a phosphorus-doped n-type (100) silicon-on-insulator (SOI) substrate (resistivity 8–12 ohm cm) patterned using standard photolithography. A 25 nm silicon dioxide layer covers the channel region to passivate the surface states, which act as carrier generation-recombination centres. The devices have a simple back-gated device configuration on a silicon substrate, where underlying buried oxide was used as the back gate. The drain current (I_d) from the source to the drain is controlled by the gate voltage (V_g) from the substrate through the buried oxide of 90 nm thickness. The electrodes attached along with the channel (Fig. 1b) for measuring the semiconductor resistivity by the four point probe technique were not employed in this work. Figure 1c shows an atomic force microscope (AFM) image of single-ion incident sites in a fission track detector, fabricated by SII.

Typical electrical characteristics (Fig. 2) show that I_d increases as source-to-drain voltage (V_d) increases at a positive back-gate bias, and saturates beyond the pinch-off point. Figure 2a confirms that our device is based on the standard field-effect transistor (FET) theory, and that the Fermi level in the channel is controlled by the back-gate and drain bias. The device exhibits an accumulation-mode n-channel transistor behaviour. In advance of SII, the initial threshold voltage (V_{th}^i) was evaluated by the extrapolation of the linear part of the curve to the V_g axis in the V_g dependence of I_d under a certain V_d of 0.1 V (Fig. 2b). The measurement was carried out at room temperature by using a semiconductor parameter analyser (Keithley 4200-SCS) and in a vacuum to avoid the disturbance caused by adsorbates.

Doubly charged P single ions were implanted one-by-one at 30 kV into the channel region at a centre-to-centre distance (pitch) of 100 nm through the 25-nm-thick surface oxide by SII, as shown in Fig. 3a, where a part of the channel region (0.3 μ m $\times$ 3.2 μ m) is indicated. The number of implanted ions in the channel was set to be 96. The projected range and the projected range straggling were calculated to be 86 nm and 22 nm, respectively, by using Monte Carlo code SRIM2003 (<http://www.SRIM.org>). For comparison with the

¹Consolidated Research Institute for Advanced Science and Medical Care (ASMeW), ²School of Science and Engineering, Waseda University, 513 Wasedatsurumaki-cho, Shinjuku-ku, Tokyo 162-0041, Japan.

ordered array, samples with random doping were also prepared by implanting P ions at an interval of 300 nm under the condition that the aiming precision was intentionally lowered down to 170 nm (Fig. 3b). The number of ions was set to be exactly the same as that of the ordered doping. The implanted samples were then lamp-annealed at a temperature of 900 °C for 3 min in N₂ to electrically activate the implanted ions. V_{th} measurement was carried out under the same condition as the initial V_{th} measurement and the difference between V_{th} before and after the SII was evaluated.

The V_{th} shift of our devices is now derived by using standard FET equations. The V_d dependence of I_d in a device with channel length L , width W and thickness t_{SOI} in the low-bias linear region is deduced as $I_d = (Wt_{SOI}/L)\mu_n C_T (V_g - V_{th})V_d$, where $V_{th} = (1/C_T)(Q_{it} - qN_D - qN_D^i) = V_{th}^i - qN_D/C_T$. Here μ is the

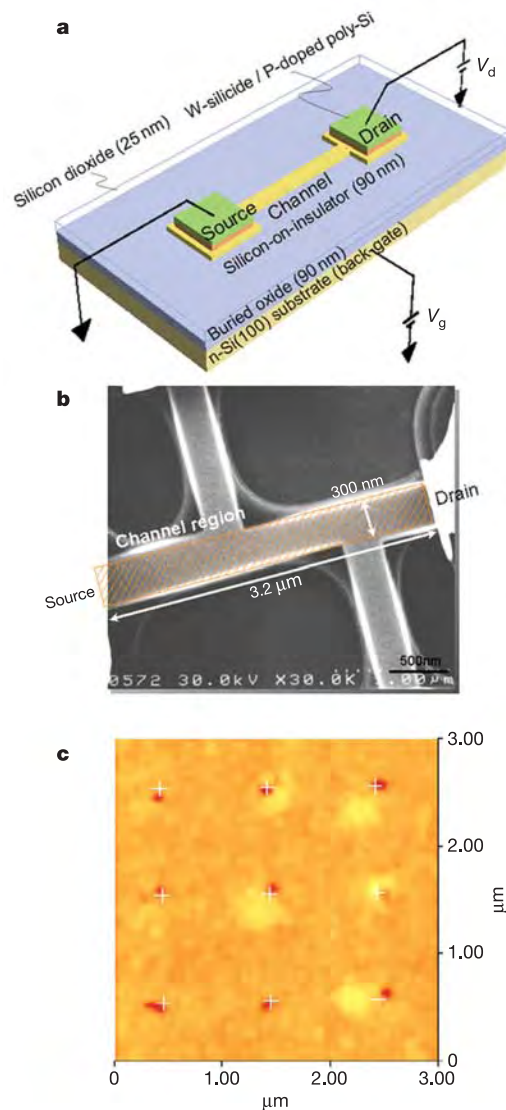


Figure 1 | Experimental device and results of single ion implantation. **a**, Illustration of device structure for controlling V_{th} through the control of dopant atom number and position by single ion implantation (SII). The oxide film (which is not displayed) covers the channel region to passivate the surface states. The current from source to drain is controlled by the gate voltage (V_g) from the back through the buried oxide. V_d , drain voltage. **b**, A scanning electron microscope (SEM) image of a representative device. The feature size is 300 nm wide, 3.2 μ m long and 90 nm thick. Black scale bar at bottom right, 500 nm. **c**, Typical atomic force microscope (AFM) image of etch-pits created in a fission track detector by single ions. The white crosses indicate the aimed positions.

mobility, q the elementary charge, N_D^i the initial dopant concentration, N_D the concentration of implanted ions, C_T the total capacitance per unit area, and Q_{it} the Si/SiO₂ interface-trapped charges. V_{th} monotonically decreases as the number of implanted ions increases. We define the difference between the V_{th} shift (ΔV_{th}) before and after single ion implantation as $\Delta V_{th} = V_{th} - V_{th}^i = -qN_D/C_T$. ΔV_{th} is directly proportional to the number of newly added dopants, and hence is a good means of counting the number of added dopants. By assessing ΔV_{th} , we eliminate the fluctuations unintentionally incorporated during the device fabrication processes, such as the fluctuation in initial dopant distribution, lithographical channel size and film thickness.

Figure 4 shows the histograms of ΔV_{th} distribution obtained from 10 FETs with ordered channel doping and 10 FETs with random channel doping. The ΔV_{th} distribution of FETs with the ordered dopant array (Fig. 4a) is much narrower than that of FETs with random channel doping (Fig. 4b). The ΔV_{th} values of FETs with random implantation deviated considerably from the mean value (Fig. 4b); the fluctuation in the dopant atom position from one device to another explains this dispersion. Gaussian fitting in the ordered dopant distribution shows a standard deviation of only 0.1 V, which is three times smaller than the random dopant distribution. We attribute the reduction of V_{th} fluctuation to the precise control of both dopant atom number and position. In addition, we find a pronounced difference in that the average value of ΔV_{th} (−0.4 V) for the ordered dopants is two times lower than that (−0.2 V) of the FETs with the random distribution of dopants, even though the dopant number is exactly same for both FETs. The larger negative ΔV_{th} for the ordered distribution indicates that the channel is open under the lower gate voltage.

The theoretical value of ΔV_{th} is calculated by using equation

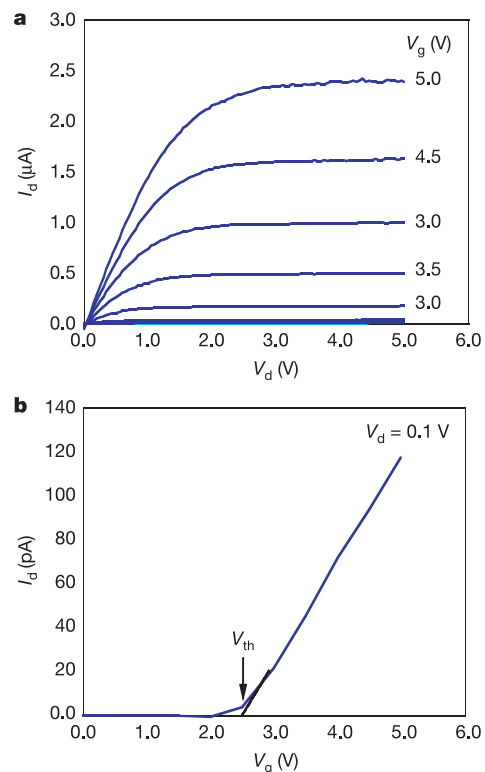


Figure 2 | Typical electrical characteristics. **a**, Drain current (I_d) versus drain voltage (V_d) characteristic at increasing gate voltage (V_g) in steps of 0.5 V starting from 0 V and going up to 5 V. The data indicate the standard electric field effect transistor (FET) behaviour. **b**, V_g dependence of I_d at a grounded source and a small drain potential of 0.1 V. Linear extrapolation gives a threshold voltage (V_{th}).

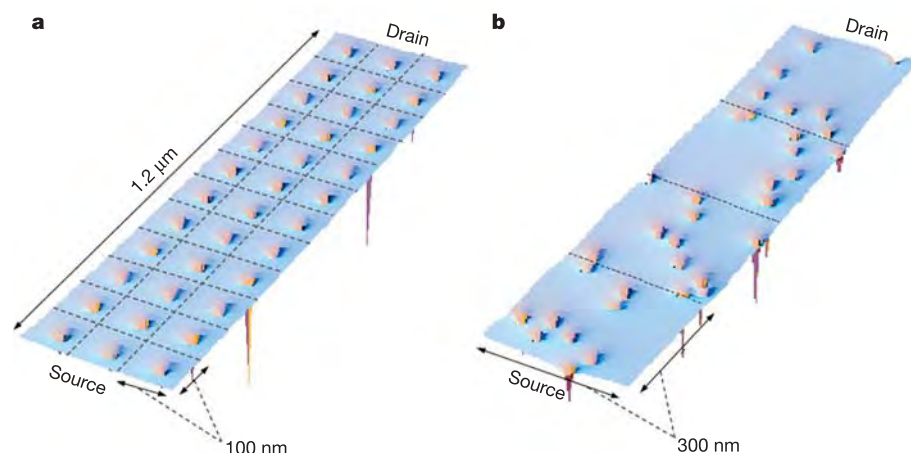


Figure 3 | Calculated potential distribution of the channel region. **a**, Potential distribution of the ordered dopant array. The energy is measured upwards. Phosphorus single ions were implanted at a pitch of 100 nm by SII. **b**, Potential distribution of the conventional random

distribution of dopant atoms, made by implanting ions at an interval of 300 nm under the condition of a lowered aiming accuracy of 170 nm, shown for comparison.

$\Delta V_{th} = -qN_D/C_T$ in which the individual dopant distribution is, *a priori*, not taken into account. The total capacitance of our sample is expressed as a series combination of the buried oxide capacitance (C_{BOX}) and the semiconductor substrate capacitance (C_s), and calculated to be $C_T = C_s C_{BOX} / (C_s + C_{BOX}) = 6.7 \times 10^{-17}$ F.

When 96 ions are implanted, the ΔV_{th} is calculated to be -0.2 V, and this value coincides with the average ΔV_{th} obtained in random channel doping. The larger negative shift of ΔV_{th} (-0.4 V) in the ordered sample cannot be explained by the conventional model, where the discrete nature of the dopant atom is ignored. To understand the larger negative shift of ΔV_{th} from the viewpoint of the discrete dopant model, we calculate the distribution of Coulomb potentials produced by ionized dopant atoms in the channel region. Figure 4 exhibits the contour map of Coulomb potentials. Here no bias is applied to the source-to-drain and the gate electrode. The potential is assumed to be Coulombic, $q^2/(4\pi\epsilon_s r)$, without the screening effect by mobile carriers and the electron energy is measured upwards (here ϵ_s and r represent respectively semiconductor permittivity and distance from the centre of the atom).

The lighter regions in Fig. 4c and d correspond to the potential barrier for electrons injected from the source. When we apply a positive voltage to the gate at a particular drain voltage, a conductive path is formed from source to drain in the channel region at a certain gate voltage, which corresponds to the threshold voltage of the device. The ordered dopant array forms a homogeneous potential distribution in the channel, as shown in Fig. 4c, resulting in the formation of a uniform current path.

In Fig. 4d, dopant atoms are randomly located within the channel region, so there is considerable variance in electrostatic potential at any point in the device. The current in the presence of potential fluctuations percolates through the 'valleys' in the potential landscape. The location and magnitude of the potential 'valleys' will strongly depend on the arrangement of the dopant ions. The current path through the channel region differs from device to device, leading to deviation in V_{th} . In the random channel doping, some parts of the channel have already reached the conductive state (darker areas), whereas others are still in the non-conductive state (lighter areas). Thus, the remaining non-conductive regions are due to the nonhomogeneous potential block in the current path formation.

As shown in Fig. 4c, the uniform channel potential in the ordered sample is obviously lower than that in the random sample (Fig. 4d). We conclude that the uniformity of channel potential lowers the voltage required to open the channel from source to drain, which allows for early turn-on in parts of the channel and results in the lower threshold voltage. Thus marked improvements in device properties could be obtained for future semiconductor electronics if dopant distributions could be precisely controlled. How a similar scheme could be implemented in mass production remains an open question.

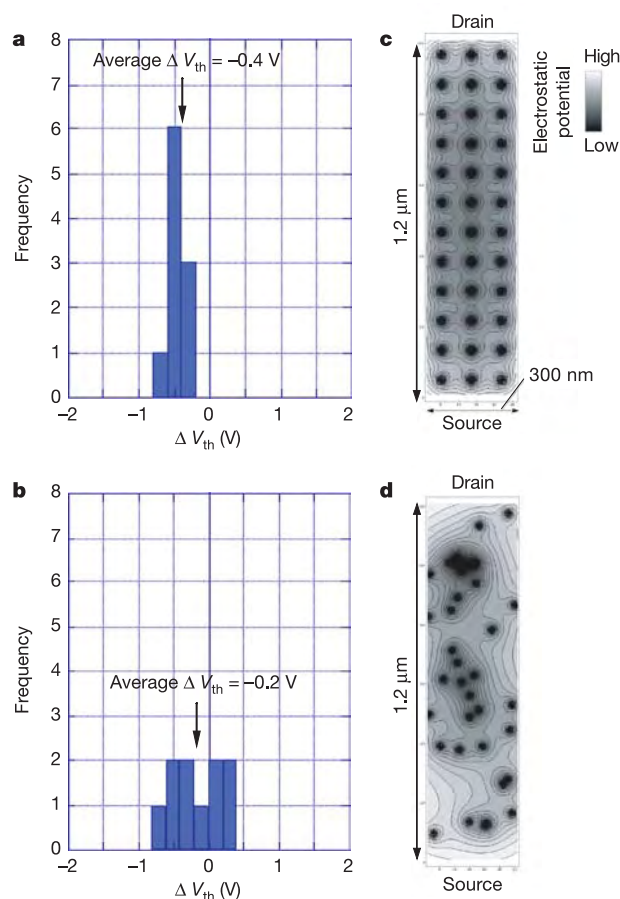


Figure 4 | Histograms of V_{th} shift (ΔV_{th}) before and after single-ion implantation from 10 resistors. **a**, Ordered dopant distribution. **b**, Conventional random dopant distribution. Gaussian fitting (curve) in the ordered dopant distribution shows a standard deviation of only 0.1 V, which is three times smaller than the random dopant distribution. **c**, **d**, The contour map of the Coulomb potential in the channel with ordered (**c**) and random (**d**) dopant distribution.

Received 1 May; accepted 27 July 2005.

1. Keyes, R. W. The effect of randomness in the distribution of impurity atoms on FET thresholds. *Appl. Phys.* **8**, 251–259 (1975).
2. Mizuno, T., Okamura, J. & Toriumi, A. Experimental—study of threshold voltage fluctuation due to statistical variation of channel dopant number in MOSFET. *IEEE Trans. Electron Devices* **41**, 2216–2221 (1994).
3. Wong, H. S. & Taur, Y. Discrete dopant distribution effects in nanometer-scale MOSFETs. *Microelectron. Reliab.* **38**, 1447–1456 (1998).
4. Sano, N. & Tomizawa, M. Random dopant model for three-dimensional drift-diffusion simulations in metal-oxide-semiconductor field-effect-transistors. *Appl. Phys. Lett.* **79**, 2267–2269 (2001).
5. Ebert, Ph., Jager, N. D., Urban, K. & Weber, E. R. Nanoscale fluctuations in the distribution of dopant atoms: Dopant-induced dots and roughness of electronic interfaces. *J. Vac. Sci. Technol. B* **22**, 2018–2025 (2004).
6. Ohdomari, I. in *Proc. 1st Int. Symp. Control of Semiconductor Interfaces* (eds Ohdomari, I., Oshima, M. & Hiraki, A.) 223–240 (North-Holland, Amsterdam, 1994).
7. Matsukawa, T. *et al.* Development of single-ion implantation—Controllability of implanted ion number. *Appl. Surf. Sci.* **117/118**, 677–683 (1997).
8. Shinada, T., Ishikawa, A., Hinoshita, C., Koh, M. & Ohdomari, I. Reduction of fluctuation in semiconductor conductivity by one-by-one ion implantation of dopant atoms. *Jpn J. Appl. Phys.* **39**, L265–L268 (2000).
9. Shinada, T., Koyama, H., Hinoshita, C., Imamura, K. & Ohdomari, I. Improvement of focused ion-beam optics in single-ion implantation for higher aiming precision of one-by-one doping of impurity atoms into nano-scale semiconductor devices. *Jpn. J. Appl. Phys.* **41**, L287–L290 (2002).
10. Kane, B. E. A silicon-based nuclear spin quantum computer. *Nature* **393**, 133–137 (1998).
11. Sze, S. M. *Physics of Semiconductor Devices: Physics and Technology* 2nd edn (Wiley, New York, 2001).

Acknowledgements This work was partly supported by the Special Coordination Funds for Promoting Science and Technology, Center of Excellence (COE) research programme and the 21st-century Center of Excellence (21COE) programme from the Ministry of Education, Culture, Sports, Science and Technology, Japan. We thank Y. Shinoda for his critical reading of the manuscript.

Author Contributions The experiment was planned by T.S. Samples were prepared, and their electrical properties measured, by T.S., S.O. and T.K. SII was invented by I.O.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.S. (shina@waseda.jp).

LETTERS

Tunable gold catalysts for selective hydrocarbon oxidation under mild conditions

Mathew D. Hughes¹, Yi-Jun Xu¹, Patrick Jenkins¹, Paul McMorn¹, Philip Landon¹, Dan I. Enache¹, Albert F. Carley¹, Gary A. Attard¹, Graham J. Hutchings¹, Frank King², E. Hugh Stitt², Peter Johnston³, Ken Griffin³ & Christopher J. Kiely⁴

Oxidation is an important method for the synthesis of chemical intermediates in the manufacture of high-tonnage commodities, high-value fine chemicals, agrochemicals and pharmaceuticals; but oxidations are often inefficient¹. The introduction of catalytic systems using oxygen from air is preferred for 'green' processing². Gold catalysis is now showing potential in selective redox processes^{3–6}, particularly for alcohol oxidation^{7–10} and the direct synthesis of hydrogen peroxide^{11,12}. However, a major challenge that persists is the synthesis of an epoxide by the direct electrophilic addition of oxygen to an alkene¹³. Although ethene is epoxidized efficiently using molecular oxygen with silver catalysts in a large-scale industrial process¹⁴, this is unique because higher alkenes can only be effectively epoxidized using hydrogen peroxide^{15–17}, hydroperoxides¹⁶ or stoichiometric oxygen donors. Here we show that nanocrystalline gold catalysts can provide tunable active catalysts for the oxidation of alkenes using air, with exceptionally high selectivity to partial oxidation products (~98%) and significant conversions. Our finding significantly extends the discovery by Haruta^{18,19} that nanocrystalline gold can epoxidize alkenes when hydrogen is used to activate the molecular oxygen; in our case, no sacrificial reductant is needed. We anticipate that our finding will initiate attempts to understand more fully the mechanism of oxygen activation at gold surfaces, which might lead to commercial exploitation of the high redox activity of gold nanocrystals.

Our initial experiments investigated the oxidation of cyclohexene using 1% Au/C with oxygen using polar solvents (for example, water) in a stirred autoclave reactor or a stirred non-pressurized glass reactor. Even under relatively mild conditions (60–80 °C, 4–24 h), only CO₂, formic acid and oxalic acid were formed with up to 100% cyclohexene conversion (Table 1), but no C₆ products were observed. Experiments involving the potential reaction of expected C₆ products with water as solvent under identical conditions indicated that had these been formed they would have been observed (conversions: cyclohexene oxide 5.6%, 2-cyclohexen-1-ol 0.2%, 2-cyclohexen-1-one 6.7%, *trans*-1,2-cyclohexanediol 0%, 1,6-hexanediol 23.0%). It is therefore clear that in polar solvents the oxidation of the alkene does not proceed via the epoxide. At this stage, we carried out a key experiment in the absence of solvent but in the presence of molecular oxygen, in a stirred autoclave reactor. In this case, 2-cyclohexene-1-one and 2-cyclohexen-1-ol were observed.

In view of these promising results, we investigated a range of apolar solvents and showed that with these media we could readily achieve selectivities for C₆ products in the range 60–80%, with yields of 10–30% (Table 1), provided that a small amount of an initiator was added at the start. In the absence of an initiator (either H₂O₂ or

t-butyl hydroperoxide, TBHP) no cyclohexene conversion was observed; however, we emphasize that only catalytic amounts of these initiators are required (Supplementary Table 1). In the absence of the gold catalyst, but in the presence of the catalytic amount of the radical initiator, again no conversion was observed, confirming that the nanocrystalline gold catalyst is necessary. The most significant finding with these solvents is that the C₆ product distribution is dependent on the solvent. The highest selectivity for cyclohexene oxide (50%) was observed with 1,2,3,5-tetramethylbenzene with 1% Au/C as catalyst (Table 1), and decreasing the Au loading decreased the yield of cyclohexene oxide (Table 2, Supplementary Table 1). Investigation of the reaction profile with reaction time (Supplementary Fig. 1) indicated this has to be carefully monitored to ensure the highest yields are achieved, as the epoxide is formed as a primary product and the other products are formed as a consequence of sequential oxidation. These findings suggest that the Au/C catalysts are tunable by careful selection of the reaction conditions. With toluene as solvent, 2-cyclohexen-1-ol and 2-cyclohexen-1-one were formed and no cyclohexene oxide was formed (Table 1). The use of d₈-deuteriotoluene did not lead to any discernable change in the rate of reaction, nor did we observe the incorporation of deuterium into the products. This indicates that the solvent is not involved intimately with the reaction, and the absence of a kinetic isotope effect indicates that the solvent was not acting as a sacrificial source of hydrogen in the rate-determining step, in contrast to the way in which sacrificial H₂ is required in previous studies for propene oxidation^{18,19}.

In a further set of experiments, styrene was used as substrate, using 1,2,4,5-tetramethylbenzene with 1% Au/C, and selectivities up to 97% were observed (Supplementary Table 1). To confirm the general applicability of this oxidation catalyst, experiments were conducted with *cis*-stilbene, and again selectivities for the epoxide of up to 90% were achieved (Supplementary Table 2). The observation of high selectivities for the *trans*-epoxide indicates that the two C–O bonds of the epoxide are formed sequentially with this substrate. Interestingly, we have also found recently that the same Au/C catalysts are effective for alkane oxidation; initial experiments indicate that cyclohexane can be oxidized at 70 °C to a mixture of cyclohexanone and cyclohexanol with a combined selectivity of near 100% at low conversion, and we are currently exploring these preliminary results in more detail²⁰.

We have also investigated the modification of the Au/C catalysts using Bi. The addition of Bi is well known to affect the selectivity of supported Pt catalysts²¹, but to date effective promoters have not been reported for Au catalysts. Addition of Bi markedly enhances the selectivity for C₆ oxidation products from cyclohexene (Table 2).

¹School of Chemistry, Cardiff University, Main Building, Park Place, Cardiff CF10 3AT, UK. ²Johnson Matthey Catalysts, PO Box 1, Teeside TS23 1LB, UK. ³Johnson Matthey Catalysts, Orchard Road, Royston, Herts SG8 5HE, UK. ⁴Center for Advanced Materials and Nanotechnology, Lehigh University, 5 East Packer Avenue, Bethlehem, Pennsylvania 18015, USA.

Table 1 | Effect of solvent on selective oxidation of cyclohexene using a 1% Au/carbon catalyst

Solvent	Conversion (%)	Product selectivity				$\Sigma_{\text{sel}} \text{C}_6^*$	$\Sigma_{\text{yield}} \text{C}_6^\dagger$
							
Water	100‡	0	0	0	0	0	0
Methanol	27.1‡	0	0	0	0	0	0
THF	5.8‡	0	0	0	0	0	0
Hexane	26.1	Tr	Tr	Tr	0	0	0
Toluene	29.1	Tr	35.1	25.1	0	60.2	17.5
1,4-Dimethylbenzene	53.5	0	12	0	43.5	55.5	29.7
1,3,5-Trimethylbenzene	8	Tr	78.1	Tr	0	78.1	6.2
1,2,3,5-TMBS	29.7	50.2	26.3	0	0	76.3	22.7
1,2,4,5-TMBS/1,4-dimethylbenzene	23.1	26	42	9.1	0	77.1	17.8
Quinoline	33.2	0	10.5	0	0	10.5	3.5
1,4-Difluorobenzene	29.1	0	47.1	26.8	0	73.9	21.5
Hexafluorobenzene	15.8	8.9	36.1	2.5	0	47.5	7.5

Reaction conditions: 1% Au/C (0.22 g), C_6H_{10} (0.012 mol), 80 °C, 24 h, solvent (20 ml). For apolar solvents, *t*-butyl hydroperoxide, TBHP (5 mol% based on C_6H_{10}) was added.

* Total selectivity to C_6 partial oxidation products.

† Total yield of C_6 partial oxidation products.

‡ Products exclusively CO_2 , HCOOH and oxalic acid.

§ Tetramethylbenzene.

With the Bi-modified 0.5% Au/C catalyst, the selectivity for C_6 products was the highest observed (97.9%). The yields observed with these modified catalysts are higher than those previously reported for the oxidation of alkenes using molecular oxygen^{18,19}, where H_2 was used as a sacrificial reductant. The promotion effect was also observed with *p*-xylene as solvent (Table 2). X-ray photoelectron spectroscopic (XPS) analysis of the catalyst before and after reaction revealed a loss of Bi from the catalyst surface (Supplementary Fig. 2). The initial molar ratio of Bi: Au for the fresh sample was 0.40, which decreased to 0.25 after reaction. To rule out the possibility that the leached Bi was promoting a homogeneous reaction, the used catalyst was re-tested with a new reactant mixture—the conversion and selectivity observed were almost identical to those obtained during the initial experiment with the fresh Bi-modified Au/C catalyst (Table 2), and the Bi: Au ratio decreased further to 0.13. In both cases, the solution recovered from the catalytic reactor exhibited no catalytic activity, confirming that the Bi-modification was due to a heterogeneously catalysed reaction. Clearly, only relatively small amounts of Bi are required to enhance the selectivity of the gold catalyst. Furthermore, initial experiments using Sn, Sb and Pb as promoters (Supplementary

Table 3) show that some enhancement in activity can be observed with these dopants, and this indicates there is potential for the design of improved gold oxidation catalysts involving promoters.

We have used cyclic voltammetry to study the Bi-modified catalysts. Figure 1a shows the first voltammetric cycle of the Bi-modified 0.5% Au/C catalysts as a function of increasing bismuth coverage. The initial stages of bismuth adsorption give rise to a series of redox peaks between 0.1 and 0.6 V associated with an irreversibly adsorbed surface bismuth layer in contact with the gold surface. Such surface intermediates have already been reported for bismuth on platinum^{21,22}. The formation of the initial monolayer coverage is associated with a feature at ~0.4–0.45 V (Fig. 1a). Once these states are filled, a multilayer bismuth stripping peak is observed on the positive potential sweep at 0.2 V which grows continuously with increasing bismuth loading. Figure 1b shows the cyclic voltammograms (CVs) of the modified catalyst before and after reaction. It is evident from these data that the Bi multilayer peak intensity is much reduced after reaction, with some attenuation of the first bismuth monolayer states. This finding is in accordance with the XPS results measured pre- and post-reaction (Supplementary Fig. 2). Evidently only the adsorbed bismuth in the first monolayer is stable (and

Table 2 | Cyclohexene oxidation with molecular oxygen using unmodified and Bi-modified Au/C catalysts

Catalyst	Conversion (%)	Product selectivity			$\Sigma_{\text{sel}} \text{C}_6^*$	$\Sigma_{\text{yield}} \text{C}_6^\dagger$
						
1.0% Au/C (10 ml Bi soln)‡	23.2	42.4	50.0	1.0	93.4	21.7
1.0% Au/C (5 ml Bi soln)‡	22.5	40.8	41.8	3.2	85.8	19.3
1.0% Au/C (unmodified)‡	29.7	50.2	26.3	Tr	76.5	22.7
0.5% Au/C (10 ml Bi soln)‡	24.2	49.6	44.2	4.1	97.9	23.7
0.5% Au/C (5 ml Bi soln)‡	19.5	48.6	41.7	3.7	94.0	18.3
0.5% Au/C (1 ml Bi soln)‡	20.4	41.2	42.2	4.4	87.8	17.9
0.5% Au/C (unmodified)‡	36.4	28.6	26.1	3.0	57.7	21.0
1.0% Au/C (unmodified)§,	53.5	0	12.0	0	55.5	29.7
1.0% Au/C (10 ml Bi soln) first use§	13.9	26.6	30.9	8.6	66.1	9.2
1.0% Au/C (10 ml Bi soln) re-used§	18.7	21.4	34.2	7.0	62.6	11.7

* Total selectivity to C_6 partial oxidation products.

† Total yield of C_6 partial oxidation products.

‡ Reaction conditions: 0.22 g catalyst, C_6H_{10} (0.012 mol), 1,2,3,5-tetramethylbenzene (20 ml), TBHP (5 mol% based on C_6H_{10}), 80 °C, 24 h.

§ Reaction conditions: as ‡ but with 1,4-dimethylbenzene as solvent instead of 1,2,3,5-tetramethylbenzene.

|| Diketone formed with selectivity of 43.2%.

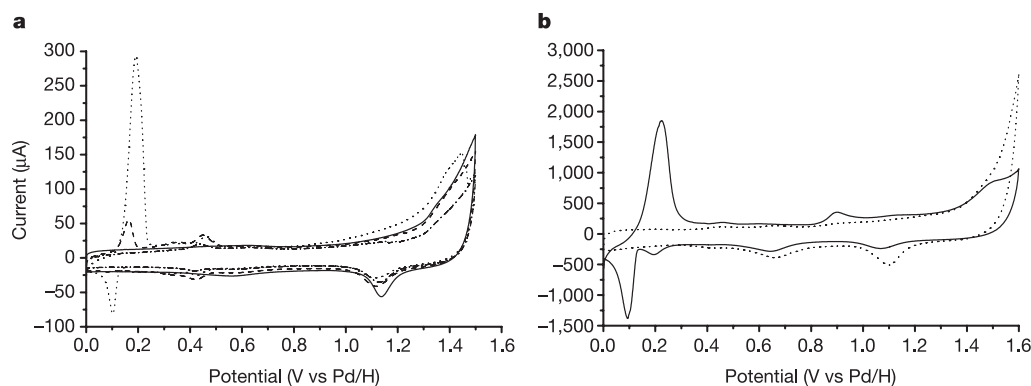


Figure 1 | Cyclic voltammograms of 0.5 wt% Au/carbon catalyst. **a**, Changes in voltammetric response as a function of irreversibly adsorbed Bi. The feature at 0.4–0.5 V is associated with Bi in the first monolayer. Multi-layers of Bi are associated with a stripping peak at 0.2 V. Bi loading (mmol): solid line, 0; dot-dashed, 0.3; dashed, 1.48; dotted, 5.92. **b**, Loss of

multi-layer Bi after reaction as demonstrated by the attenuation of the Bi stripping peak. The presence of the Bi monolayer feature suggests that some Bi remains on the catalyst. Solid line, before reaction; dotted line, after reaction.

active) under reaction conditions. However, in addition there are clearly some molecular fragments still strongly adsorbed on the surface, as indicated by the attenuation of the 'oxide' region between 1.0 and 1.6 V relative to the CV for the modified catalyst before reaction.

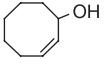
In a final set of experiments, we investigated the oxidation of *cis*-cyclooctene both in the presence (Supplementary Table 3) and absence (Table 3) of solvents. In particular, we wished to determine if the amount of radical initiator could be markedly reduced compared to our initial studies (Tables 1, 2 and Supplementary Tables 1, 2). We found that with *cis*-cyclooctene (Table 3) as substrate, high selectivities can be achieved even when the amount of TBHP is decreased to as low as 0.002 g per 10 ml substrate, representing a cyclooctene:TBHP molar ratio of 300 and a molar ratio of epoxide to peroxide of 32. With our non-optimized system, yields of 35 mol product per mol Au per h (1.9 mol per kg catalyst per h) can be achieved, which is higher than previous reports for propene epoxidation by Haruta and co-workers^{18,19}. We also found that, with careful tuning of catalyst and conditions, selective oxidation could be achieved without the need for the addition of solvent (Table 3), which is a major tenet of 'green' chemistry.

The high selectivity for the epoxide, particularly for cyclohexene and *cis*-cyclooctene, demonstrates that direct oxidation of the carbon = carbon double bond is occurring with this catalyst system. Commenting on the mechanism, we consider that HOOH and BuOOH act as initiators of a chain reaction, which is sustained by molecular oxygen. The observation of *trans*-stilbene oxide strongly suggests that for this substrate the two C–O bonds in the epoxide are not formed simultaneously; the formation of the *trans*-epoxide is evidence of a rotation about the C–C bond following the formation of the initial C–O bond. However, this behaviour may be restricted to

aromatic substrates, which may interact differently with the gold surface as compared with non-aromatic alkenes. Experiments at low conversion indicate that the epoxide is formed initially (Supplementary Fig. 1) and that the other reaction products observed represent the products of a sequential oxidation process that competes with non-selective oxidation. The active species and intermediates leading to the desired epoxide product, as well as the sequential products, are likely to be closely associated with the gold surface, and we consider that they do not involve very reactive O-centred radicals, as the preferred solvents contain benzylic C–H bonds and these are not attacked (as evident from the absence of deuterated products from the reaction in deuterotoluene). Such attack at the benzylic C–H would moreover be expected to lead to benzyl dimer formation, and the formation of selective oxidation products would be diminished because of inhibition of the chain process involving molecular oxygen; also, we see no dimeric products. The solvent does, however, influence the selectivity observed, possibly by controlling the rates of the consecutive oxidation steps as well as the formation of the active intermediates. The mediation of the solvent in the oxidation process could be via a weak interaction with the gold surface, particularly influencing the ease of electron transfer to, or from, the gold nanoparticle to establish the active chain carrier. Such a weak interaction has been observed with the C₆₀-mediated aggregation of gold nanoparticles by Brust *et al.*²³.

At present, our results show the gold catalyst to have significant potential for selective epoxide formation rather than the competing allylic oxidation. This characteristic is promising, but we note that in order to have a practical impact at the commercial level, yields will need to be significantly improved (perhaps using continuous flow reactors) and the long-term stability and performance of the catalyst demonstrated. Work along these lines is now in progress.

Table 3 | *cis*-Cyclooctene oxidation with molecular oxygen in the absence of a solvent

Catalyst	TBHP (g)	Con.(%)	Selectivity (%)				$\Sigma_{\text{sel}} \text{C-8}$
							
1%Au/C	0.12	7.9	81.2	9.3	4.1	0.5	95.1
1%Au/C	0.02	7.1	79.2	6.8	3.0	0.5	89.5
1%Au/C	0.002	1.3	82.6	7.4	2.1	0.6	92.7
No catalyst	0.008	2.0	Trace	0.0	0.0	0.0	-
Graphite	0.008	2.3	Trace	0.0	0.0	0.0	-

Reaction conditions: 0.12 g catalyst, *cis*-cyclooctene (10 ml, 0.066 mol), 80 °C, 24 h. Con., conversion.

METHODS

Preparation of supported gold catalysts. 1 wt% gold catalysts supported on carbon were prepared as follows. The carbon support (graphite, Johnson Matthey, 113.2 g) was stirred in deionized water (1 l) for 15 min. An aqueous solution of chloroauric acid (41.94% Au, Johnson Matthey, 2.38 g in 70 ml H₂O) was added slowly dropwise over a period of 30 min. After addition, the slurry was then refluxed for 30 min and following cooling, formaldehyde was added as reducing agent. The catalyst was recovered by filtration and washed until the washings contained no chloride. The catalyst was dried for 16 h at 106 °C. This method was also used to prepare 0.25 wt% Au/C and 0.5 wt% Au/C catalyst using smaller amounts of chloroauric acid. Analysis by transmission electron microscopy of the catalysts showed they all comprised Au nanocrystals with a broad size range (~5–50 nm), with a mean particle diameter of ~25 nm of multiply twinned nanocrystals (Supplementary Fig. 3), and the particle density varied with gold loading in the expected manner.

Catalyst testing. Reactions in a stirred autoclave reactor were carried out by suspending the catalyst (0.22 g) in a solution of alkene (0.012 mol) in solvent (20 ml) in a Parr autoclave (50 ml). The initiator was added as required. The autoclave was pressurized to the required pressure with oxygen (43 p.s.i.) and heated to the required temperature (50–80 °C). The reaction mixture was stirred (1,500 r.p.m.) for 4 h. Reactions in a stirred glass reactor were carried out using equal quantities of catalyst, alkene and solvent as above, but always in the presence of radical initiator (5 mol% alkene) in a 50 ml round-bottomed flask with reflux condenser connected. Experiments were carried out in an oil bath heated to 60–80 °C, with reaction mixtures stirred for 4–24 h.

Analysis of products in reactions using cyclohexene or *cis*-cyclooctene as substrate were carried out using a gas chromatograph (Varian Star 3400 CX) fitted with a DB-5 column and a flame ionization detector (FID). Samples were taken at the end of reactions; 200 µl of sample was added to 20 µl of standard (3-pentanone, 98%, Aldrich) and 0.5 µl of this sample was analysed. When styrene or *cis*-stilbene was used as substrate, analysis was carried out using high-pressure liquid chromatography with an ultraviolet detector. Reactant and product were separated using an APEX ODS 5 µm column (Jones Chromatography). The eluent was a 70:30 or 90:10 mixture of CH₃CN:H₂O for styrene and *cis*-stilbene, respectively. Samples were taken at the end of the reactions; 20 µl of sample was diluted in 10 ml of standard solution (biphenyl in CH₃CN, 1 mM solution) and 50 µl of this solution was analysed.

Preparation of Bi-doped catalysts. Bi(NO₃)₃ (0.0359 g) was dissolved in ultrapure water (50 ml) and stirred for 2 h. The required amount to obtain a particular surface concentration (referred to as 'ml Bi soln' in Table 2) was transferred to an evaporating basin containing Au/C (1 g), and stirred for 3 h, after which the stirring was then stopped and the solution was allowed to evaporate overnight. The catalyst was washed with 20 ml of ultrapure water, which was filtered off under vacuum. The catalyst was finally washed with ultrapure water (3 × 20 ml) and dried (100 °C, 1 h).

XPS measurements. Spectra were recorded with an ESCALAB 220 spectrometer using an achromatic AlK α source and an analyser pass energy of 100 eV.

Cyclic voltammetry measurements. The electrochemical cell and cyclic voltammetry methods were carried out as reported previously²⁴.

Received 6 May; accepted 2 September 2005.

- Sheldon, R. A. Heterogeneous catalytic oxidation and fine chemicals. *Stud. Surf. Sci. Catal.* **66**, 33–54 (1991).
- Gallezot, P. Selective oxidation with air on metal catalysts. *Catal. Today* **37**, 405–418 (1997).
- Bond, G. C. & Thompson, D. T. Catalysis by gold. *Catal. Rev. Sci. Eng.* **41**, 319–388 (1999).
- Fu, Q., Saltsburg, H. & Flytzani-Stephanopoulos, M. Active non-metallic Au and Pt species on ceria-based water-gas shift catalysts. *Science* **301**, 935–938 (2003).
- Baillie, J. E. & Hutchings, G. J. Promotion by sulfur of gold catalysts for crotyl alcohol formation from crotonaldehyde hydrogenation. *Chem. Commun.* 2152 (1999).

- Mohr, C., Hofmeister, H., Lucas, M. & Claus, P. Gold catalysts for the partial hydrogenation of acrolein. *Chemie-Ingenieur-Technik* **71**, 869–873 (1999).
- Biella, S., Prati, L. & Rossi, M. Selectivity control in the oxidation of phenylethane-1,2-diol with gold catalyst. *Inorg. Chim. Acta* **349**, 253–257 (2003).
- Biella, S., Prati, L. & Rossi, M. Selective oxidation of D-glucose on gold catalyst. *J. Catal.* **206**, 242–247 (2002).
- Carrettin, S., McMorn, P., Johnston, P., Griffin, K. & Hutchings, G. J. Selective oxidation of glycerol to glyceric acid using a gold catalyst in aqueous sodium hydroxide. *Chem. Commun.* 696–697 (2002).
- Carrettin, S. *et al.* Oxidation of glycerol using supported Pt, Pd and Au catalysts. *Phys. Chem. Chem. Phys.* **5**, 1329–1336 (2003).
- Landon, P., Collier, P. J., Papworth, A. J., Kiely, C. J. & Hutchings, G. J. Direct synthesis of hydrogen peroxide from H₂/O₂ using a gold catalyst. *Chem. Commun.* 2058–2059 (2002).
- Landon, P. *et al.* Direct synthesis of hydrogen peroxide from H₂ and O₂ using Pd and Au catalysts. *Phys. Chem. Chem. Phys.* **5**, 1917–1923 (2003).
- Monnier, J. R. The direct epoxidation of higher olefins using molecular oxygen. *Appl. Catal. A* **221**, 73–91 (2001).
- Boxhoorn, G. Process for the manufacture of silver-containing epoxidation catalysts. *European Patent Application* (Shell Internationale Research Maatschappij B.V., Neth., Appl. EP 87–201439, 19870727, 1988).
- Grigoropoulou, G. & Elings, J. A. Recent developments on the epoxidation of alkenes using hydrogen peroxide as oxidant. *Green Chem.* **5**, 1–7 (2003).
- Sheldon, R. A. & Van Vliet, M. C. A. in *Fine Chemicals through Heterogeneous Catalysis* (eds Sheldon, R. A. & Van Bekkum, H.) (Wiley-VCH, Weinheim, 2001).
- Lee, S. *et al.* Mechanistic insights into the conversion of cyclohexene to adipic acid by H₂O₂ in the presence of a TAPO-5 catalysts. *Angew. Chem. Int. Edn Engl.* **42**, 1520–1523 (2003).
- Sinha, A. K., Seelan, S., Tsubota, S. & Haruta, M. A three-dimensional mesoporous titanosilicate support for gold nanoparticles: vapor-phase epoxidation of propene with high conversion. *Angew. Chem. Int. Edn Engl.* **43**, 1546–1548 (2004).
- Sinha, A. K., Seelan, S., Tsubota, S. & Haruta, M. Catalysis by gold nanoparticles: epoxidation of propene. *Top. Catal.* **29**, 95–102 (2004).
- Xu, Y.-J. *et al.* Selective conversion of cyclohexane to cyclohexanol and cyclohexanone using a gold catalyst under mild conditions. *Catal. Lett.* **101**, 175–179 (2005).
- Clavilier, J., Feliu, J. M. & Aldaz, A. An irreversible structure sensitive adsorption step in bismuth underpotential deposition at platinum electrodes. *J. Electroanal. Chem.* **243**, 419–433 (1988).
- Fordham, P., Besson, M. & Gallezot, P. Selective oxidation with air of glyceric to hydroxypyruvic acid and tartronic to mesoxalic acid on PtBi/C catalysts. *Stud. Surf. Sci. Catal.* **108**, 429–436 (1997).
- Brust, M., Kiely, J. K., Bethell, D. & Schiffrin, D. J. C₆₀ mediated aggregation of gold nanoparticles. *J. Am. Chem. Soc.* **120**, 12367–12368 (1998).
- Attard, G. A. *et al.* Electrochemical evaluation of the morphology and enantioselectivity of Pt/graphite. *Appl. Catal. A* **222**, 393–405 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge the support of the Engineering and Physical Sciences Research Council (EPSRC) of the UK, Johnson Matthey plc (project ATHENA) and the European Union (project AURICAT). We also thank D. Bethell for discussions on the reaction mechanism.

Author Contributions M.D.H. and Y.-J. X. prepared and tested the catalysts under the supervision of P.L., D.I.E. and P.M. C.J.K. made the TEM measurements and A.F.C. the XPS measurements. P. Jenkins prepared the Bi-doped catalysts and made the CV measurements under the supervision of G.A.A. G.J.H. directed the research and wrote the paper. E.H.S., F.K., P. Johnston and K.G. provided discussions and advice on catalyst synthesis.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.J.H. (hutch@cardiff.ac.uk).

LETTERS

Thermochemical structures beneath Africa and the Pacific Ocean

Allen K. McNamara^{1†} & Shijie Zhong¹

Large low-velocity seismic anomalies have been detected in the Earth's lower mantle beneath Africa and the Pacific Ocean that are not easily explained by temperature variations alone^{1–11}. The African anomaly has been interpreted to be a northwest–southeast-trending structure^{3–5,7} with a sharp-edged linear, ridge-like morphology^{9,10}. The Pacific anomaly, on the other hand, appears to be more rounded in shape^{1–4,6,7,11}. Mantle models with heterogeneous composition have related these structures to dense thermochemical piles or superplumes^{12–19}. It has not been shown, however, that such models can lead to thermochemical structures that satisfy the geometrical constraints, as inferred from seismological observations. Here we present numerical models of thermochemical convection in a three-dimensional spherical geometry using plate velocities inferred for the past 119 million years²⁰. We show that Earth's subduction history can lead to thermochemical structures similar in shape to the observed large, lower-mantle velocity anomalies. We find that subduction history tends to focus dense material into a ridge-like pile beneath Africa and a relatively more-rounded pile under the Pacific Ocean, consistent with seismic observations.

Although many geodynamical studies have been performed to better understand the dynamical consequences of a dense layer in the Earth's mantle, only a subset of these have investigated the three-dimensional morphology of such a layer^{16–19}, and because of technical limitations, these have mainly considered only rectangular cartesian geometries. Cartesian experiments have been very important to our understanding of thermochemical structures in a fluid-dynamical context, but the restrictive box-shape of these experiments has precluded the ability to understand the effects of Earth-like conditions such as spherical geometry and plate-like boundary conditions. As a consequence, the actual shape of thermochemical structures predicted for the Earth could only be roughly inferred from experiments. For example, in laboratory experiments, the sides of the cartesian domain would serve as a proxy for static, circum-Pacific subduction; but this inference does not allow for the more-complicated geometry and time-dependent nature of plate boundaries, and it is also unlikely that this inference would hold for the more complicated subduction patterns associated with the African region.

In previous work with free-slip surface boundary conditions²¹, we concentrated on understanding the roles that spherical geometry and rheology play in the formation of thermochemical structures. It was found that temperature-dependent rheology leads to the formation of weak, dense piles that are passively swept aside by cold, downwelling material, resulting in a ubiquitous network of linear ridges of dense material. Only by additionally imposing an unlikely high intrinsic viscosity increase to the dense material can rounded superplumes of dense material form. This work suggested that rounded superplume structures (as proposed for the Pacific) and linear piles

(as proposed for Africa) are not expected to be present together within the Earth under the given model parameters. Like previous three-dimensional studies^{16–19}, these generalized fluid dynamical calculations provided insight into the dynamical character of thermochemical structures in a planet without Earth-like plate tectonics, but without using Earth's plate history, it was not possible to predict two large hot anomalies that could be confidently compared to the seismic anomalies observed beneath Africa and the Pacific.

Here we hypothesize that Earth's plate motion history plays a controlling role in the development of thermochemical structures that geometrically resemble to first order the general shape and locations of the large negative seismic anomalies in the lower mantle. Specifically, we investigate whether the resultant thermochemical structures are consistent with a northwest–southeast (NW–SE)-trending structure beneath Africa that has a ridge-like structure as inferred from raypath studies^{9,10}, and a more-rounded (less-linear) structure beneath the Pacific. To test this, thermochemical mantle models with initial layer thicknesses of 127 km, 255 km and 956 km for the bottom dense layer are investigated, all having a buoyancy ratio of 0.6 leading to density contrasts of 2–5% (depending upon parameters chosen for the non-dimensionalization factors of temperature and coefficient of thermal expansion). The calculations are carried out for 119 million years of model time and employ surface boundary conditions consistent with 11 stages of plate history²⁰ as used in similar studies on isochemical convection^{22,23}. Unless otherwise noted, initial temperature conditions (at 119 Myr ago) are radial temperature profiles derived from similar two-dimensional axisymmetric thermochemical calculations without plate boundary conditions.

The thermochemical extension²¹ of the parallel finite element code CitcomS²⁴ is used to solve the conservation equations of mass, momentum and energy in the Boussinesq approximation with constant thermodynamic properties. The details, along with formal definitions of the parameters and scalings, are found in previous work²¹. These calculations have much higher resolution than our previous three-dimensional spherical studies, and are performed on a mesh exceeding 3 million elements, divided between 24 processor-computational domains. Over 30 million tracers are used to characterize the compositional field. Improved algorithmic and technical computational abilities allow us to resolve viscosities consistent with Earth-like convective vigour²⁵, where the Rayleigh number (defined as $Ra = \alpha g \rho \Delta T h^3 / \kappa \eta_0$) is 2.7×10^8 ; here α , g , ρ , ΔT , h , κ and η_0 are respectively the coefficient of thermal expansion, acceleration of gravity, density, temperature drop across the mantle, mantle depth, thermal diffusivity and upper-mantle reference viscosity. A temperature- and depth-dependent rheology of the non-dimensional form is employed:

$$\eta(T, \mathbf{r}) = \eta_r(z) \exp[A(0.5 - T)]$$

¹Department of Physics, University of Colorado, Boulder, Colorado 80309-0390, USA. [†]Present address: Department of Geological Sciences, Arizona State University, Tempe, Arizona 85287, USA.

Here $\eta_r(z) = 1$ for $z < 663$ km and $\eta_r(z) = 0.1225z - 51.2$ for $663 \text{ km} \leq z \leq 2,867$ km, where η and z are the non-dimensional viscosity and dimensional depth, respectively. This formulation leads to a weak upper mantle, a $30 \times$ viscosity step at the upper–lower mantle boundary, and a $10 \times$ linear increase with depth to the base of the mantle. This depth-dependent structure is similar to that used by Bunge *et al.*^{22,23}, however, these calculations also include temperature-dependent viscosity. A non-dimensional activation coefficient, $A = 9.2103$, is chosen, which leads to a temperature-induced viscosity contrast of 10^4 . The models are heated from both below and internally with a non-dimensional heat production of 10. The calculations performed here allow for realistic plate scaling, the advantage being that our diffusion time is consistent with the material transit time. The use of plate velocity boundary conditions guides subduction to occur at locations determined by the plate history. We were careful to employ a convective vigour (Rayleigh number) that is consistent with the plate boundary conditions guiding rather than driving the flow, which minimizes concerns associated with employing kinematic boundary conditions.

Figure 1 shows different perspectives of the three-dimensional compositional fields of the three calculations for a time near the end of the calculation equivalent to the present day. The first two rows show the three-dimensional seismic tomography fields of Ritsema

*et al.*⁴ and Grand⁷ for comparison to the three geodynamical results displayed in the lower three rows. As observed in previous work^{17–19,21}, dense material is swept into ridges, but for the two lower volume cases (127 km, 255 km), the plate history incorporated here acts to focus and overlap several of these ridges under the Pacific, forming a large, contiguous pile of dense material. In contrast, material is piled into a NW–SE-directed ridge under Africa which curves eastward under Europe. The Pacific anomaly is areally extensive and contains multiple topographic peaks, whereas the African structure is a single-peaked ridge of mid-mantle height. The more-voluminous dense material case (956 km) develops extensive dense piles present everywhere except for areas in the immediate vicinity of subduction regions, and this case would produce anomalies much larger than those observed in tomography.

Temperature fields in both map format and equatorial slices for the final time (present day) of the three calculations with different initial layer thicknesses are shown in Fig. 2a. For comparison, two tomography maps^{3–5} are shown at lower-mantle depth slices in Fig. 2b and c. As expected, the pattern of subducted slabs is similar for all three cases in shallower mantle regions. At greater depths, the dense piles of material are easily discerned in the temperature field because of the high degree of heat they retain. The general trend of the piles is similar for all three cases, and in the two thinner layer

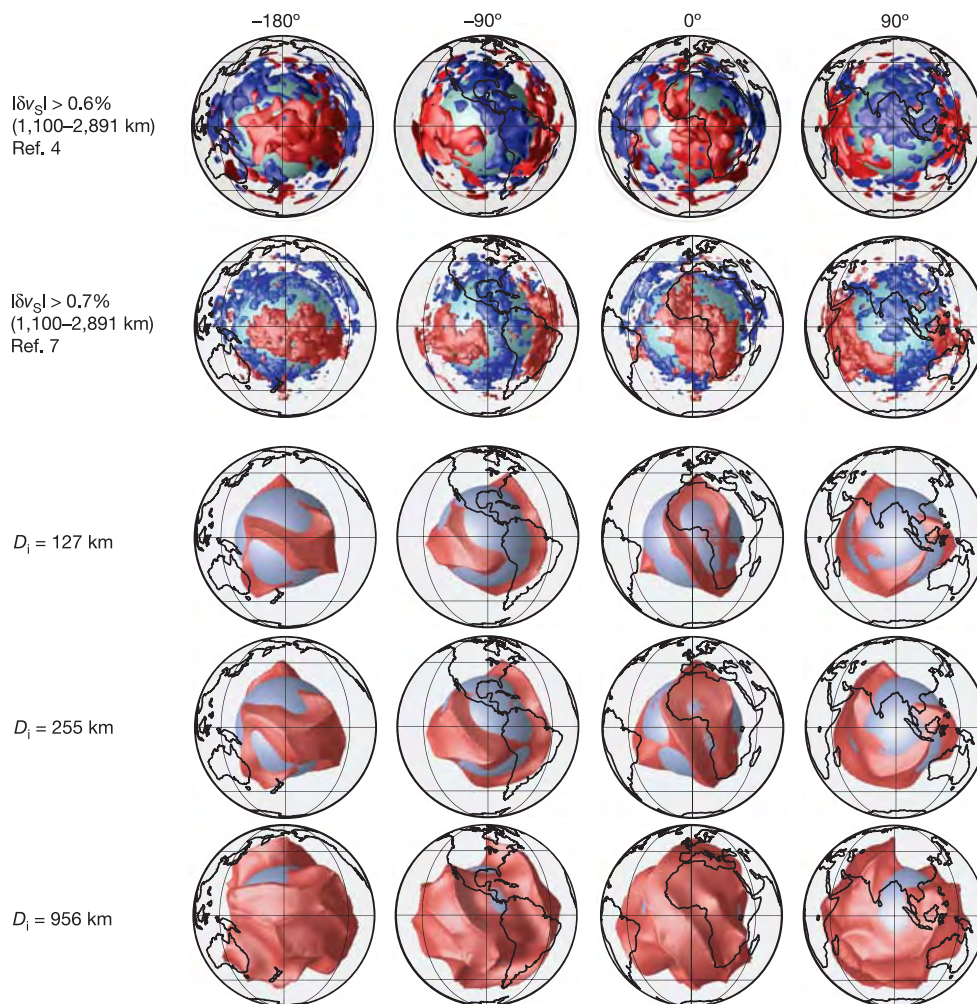


Figure 1 | Perspective views of two tomography models^{4,7} and geodynamic compositional fields at the end of the calculations, corresponding to the present day. The four columns represent hemispherical views centred on (left to right) the international dateline, longitude -90° , the prime meridian, and longitude 90° . The tomographic plots display isosurface

contours of the shear wave velocity anomaly, δv_s , and do not include data at depths less than 1,100 km to aid in visualization. Blue and red surfaces represent faster and slower than average velocities, respectively. The geodynamic results are from calculations with initial layer thicknesses (D_i) of 127 km, 255 km and 956 km.

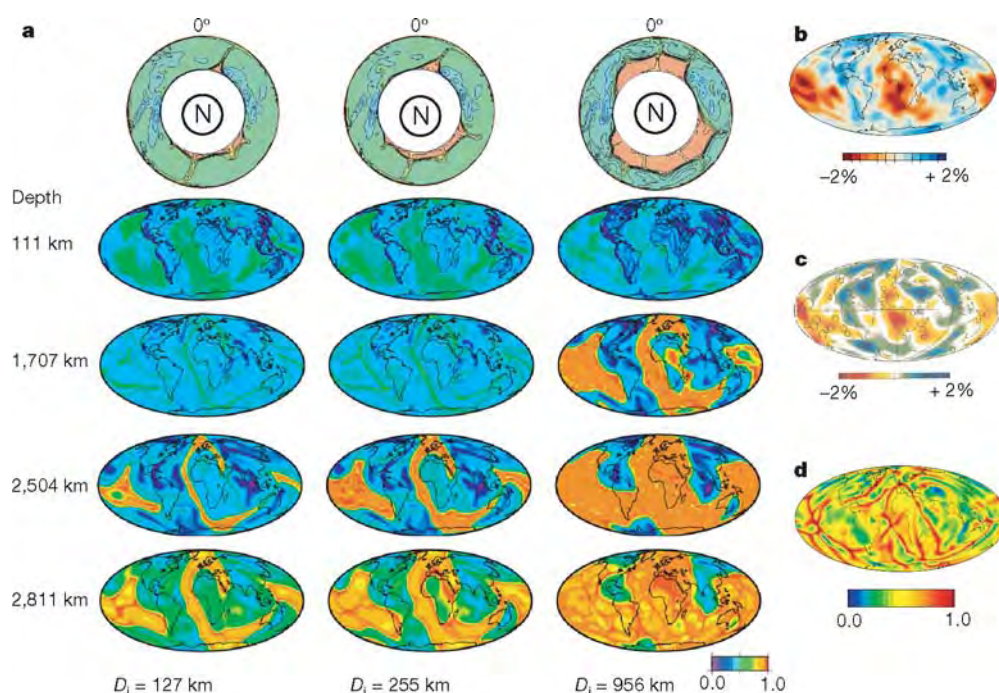


Figure 2 | Temperature and tomography maps. **a**, Present-day temperature fields for cases with initial dense layer thicknesses of 127 km, 255 km and 956 km (columns). The top row contain equatorial slices (North Pole pointing out of the page, prime meridian at the top). The other rows show temperature slices for 111 km, 1,707 km, 2,504 km and 2,811 km depth. **b**, **c**, Tomography maps at 2,811 km depth⁴ (**b**) and the core–mantle

boundary⁵ (**c**; with hotspots shown as circles) for comparison to the bottom set of geodynamical results in **a**. Temperature map (2,811 km depth) from an isochemical case lacking a dense component. The temperature and tomography maps display non-dimensional temperature and shear wave velocity anomaly, respectively.

cases, there is a NW–SE ridge structure beneath Africa that bends eastward into Europe and the Indian Ocean, not unlike that observed in tomography^{1,3,4,7}, and an intersection of ridges under the Pacific, resulting in a large continuous pile of dense material there. These thermal maps are strikingly different from those of an identical isochemical calculation (Fig. 2d), which reveal a much smaller-scale and ubiquitous structure of upwellings, similar to previous isochemical studies²³.

It is found that the volume of dense material controls the areal extent and the height of these structures, and the thickest dense layer case does not predict isolated thermochemical structures in the lower mantle. We have discovered from other calculations (not shown here) that decreasing the density contrast leads to decreased areal extent and increased height of the structures, but the general trend of the piles remains unchanged.

The role that initial condition plays in these calculations was also investigated. Whereas the previous cases started with a laterally uniform temperature field, we performed two additional thermochemical cases for the 255-km-thick initial layer case in which the

initial 119-Myr-ago plate velocity was maintained for an additional 60 and 120 Myr before starting the time sequence of plate velocity boundary conditions. Present-day results are shown in Fig. 3, and should be compared to the lowermost panel of the 255 km case in Fig. 2. With the new initial conditions, the general trend of a NW–SE ridge beneath Africa and a convergence of ridges beneath the Pacific remains, although the detailed morphology may differ somewhat. More specifically, extending the initial stage of plate motion acts to impose a pre-existing slab structure at the beginning of the calculations which effectively and artificially extends the subduction time related to the Africa–Eurasian convergence, causing excessive amounts of slab material to accumulate beneath present-day Africa. As a result of slow thermal diffusion, this material resides there until the present day, producing a somewhat different morphology.

This work investigates spherical thermochemical convection using a depth- and temperature-dependent viscosity and a realistic plate history as surface boundary conditions. A ‘first choice’ of Earth-like material parameters has led to a predicted geometry of thermochemical piles that resembles to first order the anomalies observed in seismic tomography. In addition, the sharp NW–SE-trending single-ridged African pile is consistent with that inferred from raypath studies^{9,10}. The allowable material parameter space is immense, and because the ‘first-choice’ parameters produce structures geometrically similar to those observed, we propose that the first-order morphology of a dense component is not overly sensitive to input parameters. We note that uncertainty in geodynamic parameters warrants a much broader future study, and it is not expected that current results would match higher-order geometric features that are seismically observed. Because of the imperfect distribution of seismic sources and receivers, tomography provides a filtered view of actual lower-mantle structures, and we propose that the approach used here combined with tomographic filtering of geodynamic data promises to provide a wealth of previously unavailable material information as

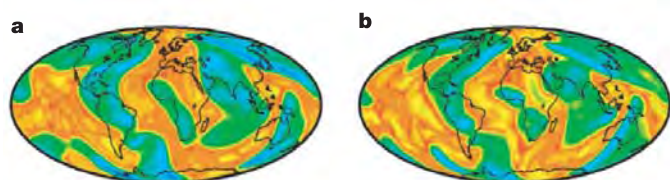


Figure 3 | Effect of initial condition. Temperature fields at 2,811 km depth for thermochemical cases with an initial dense layer thickness of 255 km. The initial stage of plate motion was held fixed for **a**, 60 Myr, and **b**, 120 Myr, before the entire time sequence of plate motion began. These should be compared with the lower panel of the 255 km case in Fig. 2. Colour scale as in Fig. 2a.

more intensive parameter searches are used to more closely relate geodynamic predictions to higher-order seismic results.

Received 13 October 2004; accepted 21 July 2005.

1. Su, W. & Dziewonski, A. M. Simultaneous inversion for 3-D variations in shear and bulk velocity in the mantle. *Phys. Earth Planet. Inter.* **100**, 135–156 (1997).
2. Breger, L. & Romanowicz, B. Three-dimensional structure at the base of the mantle beneath the central Pacific. *Science* **282**, 718–720 (1998).
3. Ritsema, J., van Heijst, H. J. & Woodhouse, J. H. Complex shear wave velocity structure imaged beneath Africa and Iceland. *Science* **286**, 1925–1928 (1999).
4. Ritsema, J., van Heijst, H. J. & Woodhouse, J. H. Global transition zone tomography. *J. Geophys. Res.* **109**, doi:10.1029/2003JB002610 (2004).
5. Kuo, B. Y., Garnero, E. J. & Lay, T. Tomographic inversion of S-SKS times for shear velocity heterogeneity in D^{''}: degree 12 and hybrid models. *J. Geophys. Res.* **105**, 28139–28157 (2000).
6. Masters, G., Laske, G., Bolton, H. & Dziewonski, A. in *Earth's Deep Interior: Mineral Physics and Tomography from the Atomic to the Global Scale* (eds Karato, S. et al.) 63–87 (Geophys. Monogr. Ser., Vol. 117, AGU, Washington DC, 2000).
7. Grand, S. P. Mantle shear-wave tomography and the fate of subducted slabs. *Phil. Trans. R. Soc. Lond. A* **360**, 2475–2491 (2002).
8. Wen, L., Silver, P., James, D. & Kuehnel, R. Seismic evidence for a thermochemical boundary at the base of the Earth's mantle. *Earth Planet. Sci. Lett.* **189**, 141–153 (2001).
9. Ni, S., Tan, E., Gurnis, M. & Helmberger, D. Sharp sides to the African superplume. *Science* **296**, 1850–1852 (2002).
10. Ni, S. & Helmberger, D. V. Seismological constraints on the South African superplume: Could be the oldest distinct structure on Earth. *Earth Planet. Sci. Lett.* **206**, 119–131 (2003).
11. Ishii, M. & Tromp, J. Constraining large-scale mantle heterogeneity using mantle and inner-core sensitive normal modes. *Phys. Earth Planet. Inter.* **146**, 113–124 (2004).
12. Tackley, P. J. in *The Core-Mantle Boundary Region* (eds Gurnis, M., Wyssession, M. E., Knittle, E. & Buffett, B. A.) 231–253 (Geodyn. Ser. Vol. 28, AGU, Washington DC, 1998).
13. Davaille, A. Simultaneous generation of hotspots and superswells by convection in a heterogeneous planetary mantle. *Nature* **402**, 756–760 (1999).
14. Kellogg, L. H., Hager, B. H. & van der Hilst, R. D. Compositional stratification in the deep mantle. *Science* **283**, 1881–1884 (1999).
15. Tackley, P. J. Mantle convection and plate tectonics: toward an integrated physical and chemical theory. *Science* **288**, 2002–2007 (2000).
16. Davaille, A., Girard, F. & Le Bars, M. How to anchor hotspots in a convecting mantle? *Earth Planet. Sci. Lett.* **203**, 621–634 (2002).
17. Jellinek, A. M. & Manga, M. The influence of a chemical boundary layer on the fixity, spacing, and lifetime of mantle plumes. *Nature* **418**, 760–763 (2002).
18. Jellinek, A. M. & Manga, M. Links between long-lived hot spots, mantle plumes, D^{''}, and plate tectonics. *Rev. Geophys.* **42**, doi:10.1029/2003RG000144 (2004).
19. Tackley, P. J. Strong heterogeneity caused by deep mantle layering. *Geochem. Geophys. Geosyst.* **3**, 1024, doi:10.1029/2001GC000167 (2002).
20. Lithgow-Bertelloni, C. & Richards, M. A. The dynamics of Cenozoic and Mesozoic plate motions. *Rev. Geophys.* **36**, 27–78 (1998).
21. McNamara, A. K. & Zhong, S. Thermochemical structures within a spherical mantle: Superplumes or piles? *J. Geophys. Res.* **109**, B07402, doi:10.1029/2003JB002847 (2004).
22. Bunge, H. P. et al. Timescales and heterogeneous structure in geodynamic Earth models. *Science* **280**, 91–95 (1998).
23. Bunge, H. P., Richards, M. A. & Baumgardner, J. R. Mantle-circulation models with sequential data assimilation: inferring present-day mantle structure from plate-motion histories. *Phil. Trans. R. Soc. Lond. A* **360**, 2545–2567 (2002).
24. Zhong, S., Zuber, M. T., Moresi, L. & Gurnis, M. Role of temperature-dependent viscosity and surface plates in spherical shell models of mantle convection. *J. Geophys. Res.* **105**, 11063–11082 (2000).
25. Mitrovica, J. X. & Forte, A. M. A new inference of mantle viscosity based upon joint inversion of convection and glacial isostatic adjustment data. *Earth Planet. Sci. Lett.* **225**, 177–189 (2004).

Acknowledgements We thank C. Lithgow-Bertelloni and T. Becker for furnishing us with the plate velocity data used in these calculations. We also thank J. Ritsema and E. Garnero for discussions and for help in generating figures for this manuscript. This work was supported by the David and Lucile Packard Foundation and the National Science Foundation.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.K.M. (allen.mcnamara@asu.edu).

LETTERS

Helium solubility in olivine and implications for high $^3\text{He}/^4\text{He}$ in ocean island basalts

Stephen W. Parman^{1†}, Mark D. Kurz², Stanley R. Hart² & Timothy L. Grove¹

High $^3\text{He}/^4\text{He}$ ratios found in ocean island basalts are the main evidence for the existence of an undegassed mantle reservoir^{1–3}. However, models of helium isotope evolution depend critically on the chemical behaviour of helium during mantle melting. It is generally assumed that helium is strongly enriched in mantle melts relative to uranium and thorium, yet estimates of helium partitioning in mantle minerals have produced conflicting results^{4–6}. Here we present experimental measurements of helium solubility in olivine at atmospheric pressure. Natural and synthetic olivines were equilibrated with a 50% helium atmosphere and analysed by crushing *in vacuo* followed by melting, and yield a minimum olivine–melt partition coefficient of 0.0025 ± 0.0005 (s.d.) and a maximum of 0.0060 ± 0.0007 (s.d.). The results indicate that helium might be more compatible than uranium and thorium during mantle melting and that high $^3\text{He}/^4\text{He}$ ratios can be preserved in depleted residues of melting. A depleted source for high $^3\text{He}/^4\text{He}$ ocean island basalts would resolve the apparent discrepancy⁷ in the relative helium concentrations of ocean island and mid-ocean-ridge basalts.

Although there is a general consensus that Earth started with an initial $^3\text{He}/^4\text{He}$ ratio of $\sim 120R_a$ ($R_a = (^3\text{He}/^4\text{He})_{\text{sample}} / (^3\text{He}/^4\text{He})_{\text{atmosphere}}$), its subsequent evolution is not well constrained⁸. In the ‘standard’ model, He is assumed to be more incompatible than U + Th (the parent isotopes of ^4He) during melting⁹. If so, any melting event will leave the mantle residue enriched in U + Th relative to He, and it will evolve radiogenic He isotope ratios (low $^3\text{He}/^4\text{He}$; Fig. 1a). In this model, the unmelted, undegassed mantle has the highest $^3\text{He}/^4\text{He}$ ratios and highest He concentrations of any mantle reservoir and is the likely source of high $^3\text{He}/^4\text{He}$ in ocean island basalt (OIB)^{2,9}.

In an alternative model, He is assumed to be more compatible than U + Th during melting, in which case melting decreases the parent/daughter ratio (U + Th)/He of the mantle. Therefore the $^3\text{He}/^4\text{He}$ ratio of a depleted residue will decrease more slowly with time than an undegassed reservoir (Fig. 1b). In this case, the highest $^3\text{He}/^4\text{He}$ mantle reservoirs are the depleted residues of melting, but unlike in the standard model, these high $^3\text{He}/^4\text{He}$ sources will have low He concentrations¹⁰. The older the depletion event and the greater the degree of melting, the larger the difference will be between a depleted and undegassed source.

Most OIBs have higher $^3\text{He}/^4\text{He}$ ratios and lower gas contents than mid-ocean-ridge basalt (MORB)^{11–13}. This has been termed the ‘helium paradox’⁷ and is more consistent with the alternative model than the standard model. However, concentrations of noble gases in lavas are dominated by shallow degassing processes, and the relative gas concentrations in OIBs and MORBs may not reflect the gas concentrations in their sources. Studies of concentration ratios (for example He/Ne) indicate that simple degassing cannot explain

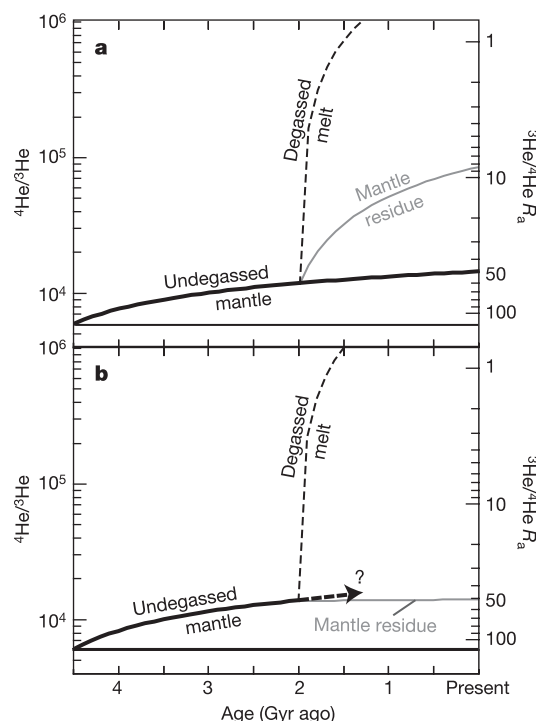


Figure 1 | Age versus He isotope ratios for two different helium isotopic evolution models. Both start with an initial $^3\text{He}/^4\text{He}$ of $120R_a$ (ref. 8). **a**, The standard model, in which helium is more incompatible than U and Th. The undegassed mantle (thick black line, $(^{238}\text{U}/^{238}\text{Th})_{\text{present}} = 450$, $(^{232}\text{Th}/^{238}\text{U})_{\text{present}} = 4$) is assumed to be sampled by OIB and so the model parameters are set to produce a $^3\text{He}/^4\text{He}$ ratio of 50 (the highest measured $^3\text{He}/^4\text{He}$ in OIB³⁰) in the undegassed mantle. A melting event at 2 Gyr is also shown, at which point the $(^{238}\text{U}/^{238}\text{Th})$ of the depleted mantle is increased by a factor of 28.6 relative to the undegassed mantle, such that it evolves to a present-day value of $8R_a$ (grey line). This is one model for the origin of the depleted MORB source. The melt has a low (U+Th)/He until it approaches the surface and degasses, increasing the $(^{238}\text{U}/^{238}\text{Th})$ 1,000-fold. Oceanic and crustal materials made from such degassed melts will rapidly evolve very low $^3\text{He}/^4\text{He}$ ratios (dashed line). The horizontal line shows a constant $^3\text{He}/^4\text{He}$ for reference. **b**, Alternative model in which He is more compatible than U+Th (the line styles are the same as in **a**). Here we assume that the highest $^3\text{He}/^4\text{He}$ OIBs are melts of depleted mantle. The undegassed mantle has $(^{238}\text{U}/^{238}\text{Th})_{\text{present}} = 600$ and $(^{232}\text{Th}/^{238}\text{U})_{\text{present}} = 4$, and melting at 2 Gyr produces a depleted residue with a $(^{238}\text{U}/^{238}\text{Th})$ one-tenth that of the undegassed mantle, which evolves to a present-day $^3\text{He}/^4\text{He}$ of 50. As in **a**, the degassed melt rapidly evolves low $^3\text{He}/^4\text{He}$.

¹Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Building 54, Cambridge, Massachusetts 02139, USA. ²Woods Hole Oceanographic Institute, Clark Building, Woods Hole, Massachusetts 02543, USA. [†]Present address: Department of Earth Sciences, University of Durham, Science Labs, Durham DH1 3LE, UK.

the differences in noble-gas concentrations or their relative abundances in most cases. A variety of more complex degassing models have been proposed to resolve the helium paradox within the framework of the standard model^{11,12,14}. Nevertheless, no direct observation indicates that the OIB source is gas-rich compared with the MORB source, as required by the standard model.

To help in distinguishing between the two models, we have measured the solubility of He in natural and synthetic olivine by using high-temperature controlled-oxygen-fugacity experiments at 1 atm pressure. The experiments were specifically designed to avoid the entrapment of melt or gas inclusions, problems that affected previous experimental studies^{4,5,15} (Fig. 2; see Methods). In particular, the olivine samples were not placed in contact with a melt phase. The partition coefficient of helium (D^{He}) between olivine and melt is calculated from the published solubility of He in basaltic melts^{16,17}. The melts used in those studies were close to (although not exactly in) equilibrium with the olivines used in our experiments. The errors introduced into our calculation of D should be negligible, because He solubility seems to vary little in basaltic compositions^{16,17}.

He concentrations in the experimental olivines were measured by noble-gas mass spectrometry by crushing *in vacuo* followed by melting of the crushed sample¹⁸. Crushing is thought to release only gas trapped in melt and gas inclusions, whereas subsequent melting releases the gas dissolved in the crystal structure. However, 50–70% of the total gas contained in our samples was released by crushing (Fig. 3). On the basis of extensive microscopic examination of the starting materials and run products, we believe that this was not caused by gas inclusions, bubbles or nanopores (see Methods). Moreover, both natural and synthetic olivine samples yielded the same gas release pattern (Fig. 3). If inclusions were dominating the He release this should not occur, because the two types of olivine should have very different inclusion populations. This indicates that some of the gas released during crushing might actually be dissolved in the olivine structure, as has been observed in some previous studies¹⁹. However, this evidence is circumstantial. Nanometre-scale

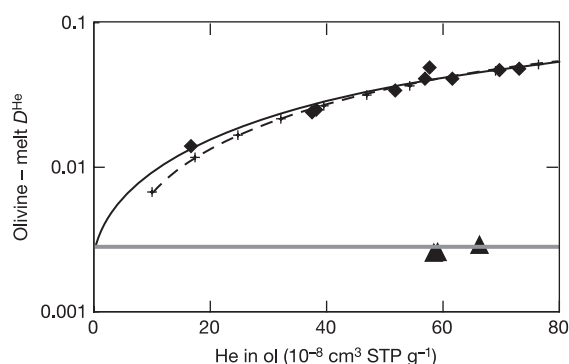


Figure 2 | Plot of helium content of experimental olivines against the calculated olivine-melt D^{He} for the data of ref. 4 (filled diamonds) and the present study (filled triangles). The data from ref. 4 show a large range of D^{He} that correlates with the He contents of the olivines. A true partition coefficient is expected to obey Henry's law and should be constant (grey horizontal line). A model in which an olivine with low He contents ($10^{-7} \text{ cm}^3 \text{ STP g}^{-1}$) is variably contaminated with the melt from the experiments fits the data very well (dashed curve) and indicates that melt inclusions in the olivine might be responsible for the changing D values and He contents in ref. 4. A linear fit (representing a mixing line) through the data (solid black curve) yields an R^2 of 0.951 and an intercept of 0.0027. The intercept represents the D^{He} value at a hypothetical zero level of contamination and is a lower limit on D^{He} in the experiments of ref. 4. The lower limit corresponds well to the D^{He} values (0.0025–0.0060) reported here and indicates that the experiments might be obeying Henry's law. The D values of this study were calculated by dividing the amount of He released by melting by the solubility of He in basaltic melt ($5 \times 10^{-4} \text{ cm}^3 \text{ STP g}^{-1} \text{ atm}^{-1}$)^{16,17}.

inclusions or voids cannot be completely ruled out as a source for the crush gas at this point.

Three additional experiments were performed in which the olivine starting material was ground to submillimetre grain sizes before the experiment, to test the influence of sintering and surface area. These samples showed slightly elevated gas release during both crushing and melting (Fig. 3), indicating that measurable quantities of gas can be trapped between grains or adsorbed on grain surfaces. This is probably the cause of high values of D for noble gases in previous studies that used fine ($<10\text{-}\mu\text{m}$ diameter) mineral powders¹⁵. However, the results show that, in the present experiments, sintering and surface adsorption were not major sources of He.

The amount of gas released by melting of the crushed powders was quite consistent, yielding a D^{He} of 0.0025 ± 0.0005 (excluding experiment 106C, see Fig. 3 caption and Methods). The similarity of D^{He} in natural, iron-bearing and synthetic, iron-free olivine indicates that vacancies might have a limited effect on He solubility because the point-defect population in olivine is primarily controlled

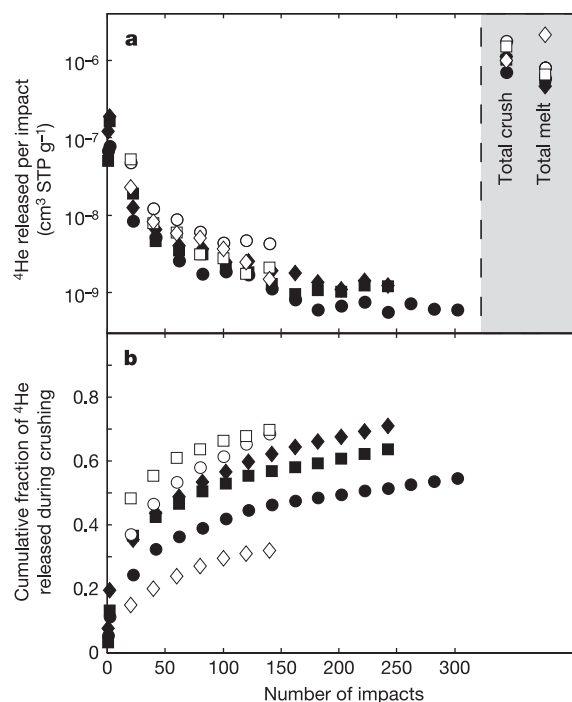


Figure 3 | Helium released from experimental olivine grains during crushing and melting. **a**, Gas released per impact. For the large-grain-size experiments 100A (filled circles, AAA gem San Carlos olivine), 102A (filled squares, AA gem San Carlos olivine) and 104A (filled diamonds, pure synthetic forsterite) and for the powder experiments 106A (open circles, $<100\text{-}\mu\text{m}$ grains, San Carlos olivine), 106B (open squares, $\sim 750\text{-}\mu\text{m}$ grains, San Carlos olivine) and 106C (open diamonds, $\sim 500\text{-}\mu\text{m}$ grains, San Carlos olivine). The values for gas released are calculated by dividing the amount of gas measured at each analytical step by the number of impacts that occurred since the last analysis and therefore represent average release values over that span of impacts. The last two symbols in each sequence (in the grey region) are the total amount of gas released by crushing and by melting. Excluding experiment 106C, the total amount of gas released by crushing is more variable ($\text{RSD} = 34\%$) than the amount released by melting ($\text{RSD} = 21\%$), indicating that some part of the crush gas might reside in a different location than the gas released by melting. **b**, Cumulative amount of gas released by crushing divided by the total amount of gas released. The symbols are the same as in **a**. Between 50% and 70% of the total (crush+melt) gas in the samples is released by crushing (excluding experiment 106C). The gas release curves flatten after ~ 150 impacts, indicating either that there might be some fraction ($\sim 20\%$) of the gas that cannot be released by crushing or that the crushing process might have reached its limit and is no longer breaking grains.

by Fe^{3+} substitutions²⁰. For the same reason, the effect of oxygen fugacity f_{O_2} on D^{He} should also be relatively small. The total amount of He released by crushing and melting (using only the large-grain-size experiments) yields an average D^{He} of 0.0060 ± 0.0007 , the maximum value that these experiments would support. Because we do not fully understand the origin of the gas released by crushing, in the following discussion we use only the lower value of 0.0025. Using the higher value would only strengthen our conclusions. This value is lower than previous experimental measurements ($0.01\text{--}0.005$)^{4,5} and within estimates of maximum D^{He} based on natural olivine–melt pairs: 0.001 (M.D.K., unpublished observations) to 0.008 (ref. 6). Note that our value for D^{He} is, strictly, valid only at the experimental conditions. We assume that D^{He} is similar at mantle melting conditions. Preliminary experiments on the effects of oxygen fugacity and temperature support this view, although the effects of pressure have yet to be quantified.

The solubility experiments indicate that ancient melt-depleted mantle might retain high $^3\text{He}/^4\text{He}$ ratios. D^{He} in olivine (ol) is similar to the estimated D^{He} for clinopyroxene (cpx)²¹, indicating that He might be distributed relatively evenly between the mantle minerals (Fig. 4). In contrast, most trace elements (including U and Th) are strongly concentrated in cpx and garnet (gt). Thus, unlike other isotopic systems (for example Sr, Nd and Pb), the parent isotopes ($^{235,238}\text{U}$ and ^{232}Th) in the He system may reside largely in a different mantle mineral from their daughter isotope (^4He). This has implications for He isotope evolution because mantle melting is non-modal. That is, the minerals do not enter the melt in the same proportion in which they are present in the mantle. Clinopyroxene and gt enter the melt at 2–3 times the rate of ol and orthopyroxene

(opx)²². So, as melting proceeds, the fraction of cpx and gt (the main carriers of the parent isotopes) in the residue will decrease and the fraction of ol (the main carrier of the daughter isotope) will increase.

During early melting of a gt or spinel (sp) lherzolite (ol+opx+cpx+sp/gt), the bulk D (the mineral–melt D values multiplied by their modal fraction in the mantle) for U and Th may be close to that of He, but will rapidly fall as cpx and gt are removed by non-modal melting (Fig. 4). In contrast, the bulk D for He will remain relatively constant or may even rise. When all of the cpx and gt are removed, leaving a harzburgite residue, $D^{\text{U+Th}}$ will be significantly lower than D^{He} (Fig. 4). For dunite, the difference will be even greater. Harzburgite and especially dunite residues should have exceedingly low (U + Th)/He, and will preserve high $^3\text{He}/^4\text{He}$ ratios relative to an undegassed source (Fig. 1b).

Given the paucity of noble-gas partitioning data, quantitative modelling assuming a particular melting style (for example batch, fractional or continuous) is premature. However, a few points can be made. If the bulk D for He is greater than the bulk D for U+Th, then both batch and fractional models will produce residues with low (U+Th)/He. Fractional melting will produce the most fractionated (lowest) (U+Th)/He, whereas batch melting will produce the least. Other melting models will fall between these two extreme cases. Combined with the effects of non-modal melting, the difference between D^{He} and $D^{\text{U+Th}}$ in olivine is so large that it eclipses most other factors (Fig. 4).

Whereas our experimental samples were specifically chosen for their lack of mineral, melt or gas inclusions, most natural minerals have such inclusions. Melt inclusions will act as phases with $D = 1$ for all elements. Gas inclusions will have high D for helium but low D for U and Th. Thus, the effective olivine–melt D^{He} in nature is likely to be higher than our measured value on ‘unnaturally’ inclusion-free minerals.

It is likely that depleted residues are present throughout the mantle. Clinopyroxene-poor lherzolites, harzburgites and dunites are produced at mid-ocean ridges²³ and subduction zones²⁴, and are subducted at convergent margins. In addition, ^{142}Nd isotope data indicate that most of the mantle might have been depleted early in Earth's history²⁵. If depleted residues can retain high $^3\text{He}/^4\text{He}$, they are potential sources for at least some of the high- $^3\text{He}/^4\text{He}$ OIB. Correlations between Sr, Nd and He isotopes have been used to support this idea²⁶. However, undegassed mantle can also yield high $^3\text{He}/^4\text{He}$ in OIB (Fig. 4). The most direct way to distinguish between the two possibilities would be to accurately constrain the pre-degassing abundances of He and other noble gases in OIB and MORB. The low He concentrations in OIB, relative to MORB, indicate that they might have ancient depleted sources, but confirming this hypothesis will require significant advances in our understanding of magmatic devolatilization mechanisms and volatile element fractionation.

METHODS

Experiments. To avoid the problems encountered in previous studies, the olivine samples were equilibrated directly with a 50% (v/v) He atmosphere in a 1-atm gas mixing furnace at 1,350 °C. The remaining 50% of the atmosphere was a mixture of CO_2 and H_2 , which was used to maintain the f_{O_2} at the Ni–NiO oxygen buffer. The samples were large (millimetre-sized) pieces of inclusion-free, gem-quality olivine. These were placed in large (centimetre-sized) capsules of San Carlos olivine and run for 17–21 days to ensure diffusive equilibrium²⁷. There was no evidence of sintering or reaction after the experiment. Three starting materials were used: two types of natural San Carlos olivine (AAA and AA grade, faceted gems) and a synthetic pure forsterite (produced by the Linde Corporation). The starting materials for experiment 100A were extensively examined for inclusions and pores before and after the experiments, by optical microscopy, back-scattered and secondary electron imaging, and annular dark-field/bright-field imaging with a scanning transmission electron microscope (STEM). The materials were inclusion-free and pore-free down to the resolution of the STEM, ~10 nm. The volume of olivine imaged with the STEM was ~10⁹ nm³.

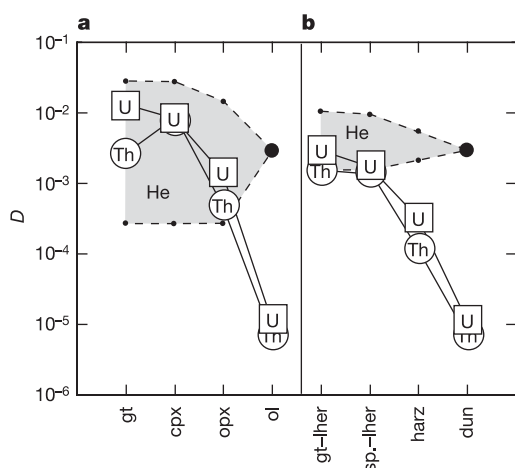


Figure 4 | Partition coefficients for U (squares), Th (circles) and He (filled circles and field) relevant to mantle melting. a, Individual mineral–melt D values for garnet (gt), clinopyroxene (cpx), orthopyroxene (opx) and olivine (ol). b, Bulk D values for garnet lherzolite (gt–lher), spinel lherzolite (sp–lher), harzburgite (harz) and dunite (dun). The D^{He} value for olivine is from this study (large filled circle). We assume that the value measured at 1 atm is applicable to melting at depth. Values for D^{He} in gt, cpx and opx have not been published. Here we use experiments on the heavier noble gases²¹ and He analyses from natural samples²⁸ to infer upper and lower bounds on D^{He} in these minerals. Partition coefficients for U and Th are averages of all published values. No partitioning data are available for spinel and so all D values are assumed to be zero. See Methods for a full discussion of how the D values were calculated and for the references used. During the early stages of melting (lherzolite assemblage), the bulk D for U and Th fall on the lower side of the D^{He} field. However, the bulk D for U and Th quickly decrease as melting proceeds and the residue approaches harzburgite (ol + opx) and dunite (ol) assemblages. The fractions of minerals [ol, opx, cpx, sp, gt] used in the calculation of the bulk D values are garnet lherzolite [0.61, 0.13, 0.13, 0.0, 0.13], spinel lherzolite [0.61, 0.20, 0.17, 0.02, 0.0], harzburgite [0.74, 0.23, 0.0, 0.03, 0] and dunite [1.0, 0, 0, 0, 0].

Three additional experiments were performed in which the starting material (non-gem-quality San Carlos olivine) was ground to <100- μm , ~500- μm and ~750- μm grain size to examine the effects of surface area and sintering. Other than the grain size, the experimental conditions were the same as for the first three experiments. The materials in run 106A, with the <100- μm grain size, sintered together and to the capsule wall, whereas those in the other two powder experiments did not. Run conditions and analytical data are provided in Supplementary Tables 1 and 2.

Gas analyses. For the large-grain-size experiments, the analyses were performed after every 20 impacts except for the first two analyses, which were taken after one impact. For the powdered experiments, analyses were taken after every 20 steps. The large-grain-size experiments (100A, 102A and 104A) show highly consistent crushing (total crush relative standard deviation (RSD) = 23%) and melting (melt RSD = 14%) release patterns. The powder experiments (106A, 106B and 106C) show noticeably higher gas release during crushing and melting. Experiment 106C in particular is anomalous with large amounts of gas released during melting and small amounts of gas released during crushing relative to the other powdered experiments. This experiment is excluded from all calculations of solubility and partitioning. Considering all of the experiments except 106C, the total amount of gas released by crushing is more variable (RSD = 34%) than the amount released by melting (RSD = 21%), indicating that some part of the crush gas might reside in a different location from the gas released by melting.

Partition coefficients. The grey field in Fig. 4 represents the range of He D values that can be supported by experimental or, lacking that, natural estimates of partition coefficients. The lower bounds on D^{He} in gt, cpx and opx are based on the published experimental D values for Ne, Ar, Kr and Xe in cpx²¹. We extrapolated the lattice-strain parabola for cpx defined by the noble gases heavier than He, which yields a value of ~0.0003. As there are no noble-gas partitioning data for opx or gt we make the *ad hoc* assumption that they have the same D^{He} as cpx. The upper bound was estimated from measurements of total He contents in ol, opx and cpx in peridotites²⁸. The data are quite scattered, with relative standard deviations up to 100%. From median values, the opx has about fivefold more He, and the cpx about tenfold more He, than the olivine crystals in the samples. For the upper curve we assume that these ratios also reflect the relative sizes of the partition coefficients. As there are no gt data for the peridotite, we assume that its D^{He} is the same as that for cpx. The peridotite data are strongly affected by inclusions, but they do indicate a plausible upper bound. The decrease in D^{He} from cpx to opx to ol in the natural data resembles the decreasing compatibility of other incompatible trace elements (for example U and Th) in these minerals (Fig. 4), and would be consistent with He partitioning being controlled by the same parameters (for example atomic size, charge and lattice-site size mismatch) that control most trace element partitioning²¹.

The mineral–melt D values for U and Th are averages of published experimental data (references in Supplementary Information). Our bulk D values calculated from the average mineral–melt D values (Fig. 4) agree well with bulk D values calculated from lattice-strain models²⁹.

Received 12 April; accepted 2 September 2005.

- Kurz, M. D., Jenkins, W. J. & Hart, S. R. Helium isotopic systematics of oceanic islands and mantle heterogeneity. *Nature* **297**, 43–47 (1982).
- Moreira, M., Doucet, S., Madureira, P. M., Lecomte, A. & Allègre, C. J. Helium–neon systematics in OIB and the nature of the source of mantle plumes. *Geochim. Cosmochim. Acta* **68**, A283 (2004).
- Allegre, C. J., Staudacher, T., Sarda, P. & Kurz, M. Constraints on the evolution of the Earth's mantle from rare-gas systematics. *Nature* **303**, 762–766 (1983).
- Hiyagon, H. & Ozima, M. Partition of noble-gases between olivine and basalt melt. *Geochim. Cosmochim. Acta* **50**, 2045–2057 (1986).
- Brooker, R. A., Heber, V., Kelley, S. P. & Wood, B. J. Noble gas partitioning behaviour during mantle melting: a possible explanation for 'the He paradox'? *EOS* **84**, Abstract V31F-03 (2003).
- Marty, B. & Lussiez, P. Constraints on rare-gas partition-coefficients from analysis of olivine glass from a picritic midocean ridge basalt. *Chem. Geol.* **106**, 1–7 (1993).
- Anderson, D. L. The helium paradoxes. *Proc. Natl Acad. Sci. USA* **95**, 4822–4827 (1998).
- Graham, D. W. in *Noble Gases in Geochemistry and Cosmochemistry* (eds Porcelli, D., Ballentine, C. J. & Wieler, R.) 247–317 (The Mineralogical Society of America, Washington DC, 2002).
- Kurz, M. D. Mantle heterogeneity beneath oceanic islands—some inferences from isotopes. *Phil. Trans. R. Soc. Lond. A* **342**, 91–103 (1993).
- Meibom, A. et al. Are high He-3/He-4 ratios in oceanic basalts an indicator of deep-mantle plume components? *Earth Planet. Sci. Lett.* **208**, 197–204 (2003).
- Yamamoto, J. & Burnard, P. G. Solubility controlled noble gas fractionation during magmatic degassing: implications for noble gas compositions of primary melts of OIB and MORB. *Geochim. Cosmochim. Acta* **69**, 727–734 (2005).
- Honda, M. & Patterson, D. B. Systematic elemental fractionation of mantle-derived helium, neon, and argon in mid-oceanic ridge glasses. *Geochim. Cosmochim. Acta* **63**, 2863–2874 (1999).
- Workman, R. K. et al. Recycled metasomatized lithosphere as the origin of the enriched mantle II (EM2) end-member: evidence from the Samoan volcanic chain. *Geochim. Geophys. Geosyst.* **5**, doi:10.1029/2003GC000623 (2004).
- Burnard, P. G., Graham, D. W. & Farley, K. A. Mechanisms of magmatic gas loss along the Southeast Indian Ridge and the Amsterdam–St. Paul Plateau. *Earth Planet. Sci. Lett.* **203**, 131–148 (2002).
- Broadhurst, C. L., Drake, M. J., Hagee, B. E. & Bernatowicz, T. J. Solubility and Partitioning of Ne, Ar, Kr, and Xe in minerals and synthetic basaltic melts. *Geochim. Cosmochim. Acta* **56**, 709–723 (1992).
- Lux, G. The behaviour of noble-gases in silicate liquids—solution, diffusion, bubbles and surface effects, with applications to natural samples. *Geochim. Cosmochim. Acta* **51**, 1549–1560 (1987).
- Jambon, A., Weber, H. & Braun, O. Solubility of He, Ne, Ar, Kr and Xe in a basalt melt in the range 1250–1600 degrees C. *Geochim. Cosmochim. Acta* **50**, 401–408 (1986).
- Kurz, M. D., Curtice, J., Lott, D. E. & Solow, A. Rapid helium isotopic variability in Mauna Kea shield lavas from the Hawaiian Scientific Drilling Project. *Geochim. Geophys. Geosyst.* **5**, doi:10.1029/2002GC000439 (2004).
- Yokochi, R., Marty, B., Pik, R. & Burnard, P. High ³He/⁴He ratios in peridotite xenoliths from SW Japan revisited: Evidence for cosmogenic ³He released by vacuum crushing. *Geochim. Geophys. Geosyst.* **6**, doi:10.1029/2004GC000836 (2005).
- Nakamura, A. & Schmalzried, H. On the nonstoichiometry and point-defects of olivine. *Phys. Chem. Miner.* **10**, 27–37 (1983).
- Brooker, R. A. et al. The 'zero charge' partitioning behaviour of noble gases during mantle melting. *Nature* **423**, 738–741 (2003).
- Baker, M. B. & Stolper, E. M. Determining the composition of high-pressure mantle melts using diamond aggregates. *Geochim. Cosmochim. Acta* **58**, 2811–2827 (1994).
- Dick, H. J. B., Fisher, R. L. & Bryan, W. B. Mineralogic variability of the uppermost mantle along mid-ocean ridges. *Earth Planet. Sci. Lett.* **69**, 88–106 (1984).
- Gill, J. B. *Orogenic Andesites and Plate Tectonics* (Springer, New York, 1981).
- Boyett, M. & Carlson, R. W. ¹⁴²Nd evidence for early (>4.53 Ga) global differentiation of the Silicate Earth. *Science* (in the press) (2005).
- Class, C. & Goldstein, S. L. Evolution of helium isotopes in the Earth's mantle. *Nature* **436**, 1107–1112 (2005).
- Trull, T. W. & Kurz, M. D. Experimental measurements of He-3 and He-4 mobility in olivine and clinopyroxene at magmatic temperatures. *Geochim. Cosmochim. Acta* **57**, 1313–1324 (1993).
- Dunai, T. J. & Baur, H. Helium, neon, and argon systematics of the European subcontinental mantle—implications for its geochemical evolution. *Geochim. Cosmochim. Acta* **59**, 2767–2783 (1995).
- Landwehr, D., Blundy, J., Chamorro-Perez, E. M., Hill, E. & Wood, B. U-series disequilibria generated by partial melting of spinel lherzolite. *Earth Planet. Sci. Lett.* **188**, 329–348 (2001).
- Stuart, F. M., Lass-Evans, S., Fitton, J. G. & Ellam, R. M. High ³He/⁴He ratios in picritic basalts from Baffin Island and the role of a mixed reservoir in mantle plumes. *Nature* **424**, 57–59 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Curtice for his long labours with the He analyses, and J. van Orman for assistance in the early planning and execution of the project. This research was supported by the NSF.

Author Contributions S.W.P. and T.L.G. performed the experiments and microscopic observations. M.D.K. performed the He analysis. S.R.H. contributed to experimental and analytical design. All authors contributed to data analysis.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.W.P. (stephen.parman@durham.ac.uk).

LETTERS

Neuroanatomy of sea spiders implies an appendicular origin of the protocerebral segment

Amy Maxmen¹, William E. Browne³, Mark Q. Martindale³ & Gonzalo Giribet²

Independent specialization of arthropod body segments has led to more than a century of debate on the homology of morphologically diverse segments^{1,2}, each defined by a lateral appendage and a ganglion of the central nervous system. The plesiomorphic composition of the arthropod head remains enigmatic because variation in segments and corresponding appendages is extreme. Within extant arthropod classes (Chelicerata, Myriapoda, Crustacea and Hexapoda—including the insects), correspondences between the appendage-bearing second (deutocerebral) and third (tritocerebral) cephalic neuromeres have been recently resolved on the basis of immunohistochemistry¹ and Hox gene expression patterns^{3,4}. However, no appendage targets the first ganglion, the protocerebrum, and the corresponding segmental identity of this anterior region remains unclear⁵. Reconstructions of stem-group arthropods indicate that the anteriormost region originally might have borne an ocular apparatus and a frontal appendage innervated by the protocerebrum⁶. However, no study of the central nervous system in extant arthropods has been able to corroborate this idea directly, although recent analyses of cephalic gene expression patterns in insects suggest a segmental status for the protocerebral region^{7–10}. Here we investigate the developmental neuroanatomy of a putative basal arthropod¹¹, the pycnogonid sea spider, with immunohistochemical techniques. We show that the first pair of appendages, the chelifores, are innervated at an anterior position on the protocerebrum. This is the first true appendage shown to be innervated by the protocerebrum, and thus pycnogonid chelifores are not positionally homologous to appendages of extant arthropods but might, in fact, be homologous to the ‘great appendages’ of certain Cambrian stem-group arthropods.

The traditional view of the arthropod head has been that the anteriormost region, containing the eyes and protocerebrum, belongs to a presegmental anterior cap, the ‘acron’^{6,12}. The term acron was developed in the context of the Articulata hypothesis in which annelids and arthropods share a recent common ancestor, with the acron being homologous with the annelid presegmental prostomium². Although the Articulata hypothesis has fallen out of favour, the acron is often cited as part of the arthropod ground plan¹². Evidence against the acron concept, drawn from onychophoran neuroanatomy¹³ and insect development, demonstrates that the arthropod protocerebral neuromere may be truly segmental^{7,9} and involved with the coordination of motor/sensory behaviour associated with appendages¹⁴.

Pycnogonids are most commonly placed as a sister taxon to the chelicerates^{15,16} or as a separate class, basal to all remaining extant arthropods¹¹ (Fig. 1). The fossil record indicates that pycnogonids branched off early during the emergence of stem-group arthropods with earliest larvae identified from the Upper Cambrian¹⁷ (about

490 Myr ago) and a crown-group (or near-crown-group) adult from the Silurian period¹⁸ (about 425 Myr ago). The antiquity of the lineage is congruent with the observation that pycnogonids share traits with other ecdysozoans (tardigrades, nematodes, priapulans and their kind), including a terminal mouth¹⁹ and a triradiate pharynx²⁰, that could be found in the arthropod ground pattern.

Because of the critical position of pycnogonids as putative basal arthropods, the structure and development of the nervous system of a Hawaiian species belonging to the genus *Anoplodactylus* were examined by immunohistochemistry in conjunction with confocal laser-scanning microscopy. The scaffolding and segmental nature of the larval central nervous system (CNS) can be clearly identified before the fusion of cephalic ganglia. The *Anoplodactylus* sp. protonymphon larva (Fig. 2a) is microscopic (about 0.05 mm) when it hatches from eggs carried exclusively by the male. The protonymphon bears a pair of functional chelifores (claw-like appendages) followed by two postcheliforal appendages²¹. The digestive tract is initially incomplete and consists of a short proboscis with a terminal mouth and a highly innervated, muscularized oesophagus. After the second protonymphon stage, the *Anoplodactylus* sp. larva encysts within a hydroid until it emerges as a juvenile. Protonymphon development involves a series of transformations during which the larval chelifores transform into adult chelifores, and the second and third larval appendages are transformed into palps and ovigers, respectively²¹. The more posterior walking legs are added in a succession of moults, beginning with limb buds forming the fourth

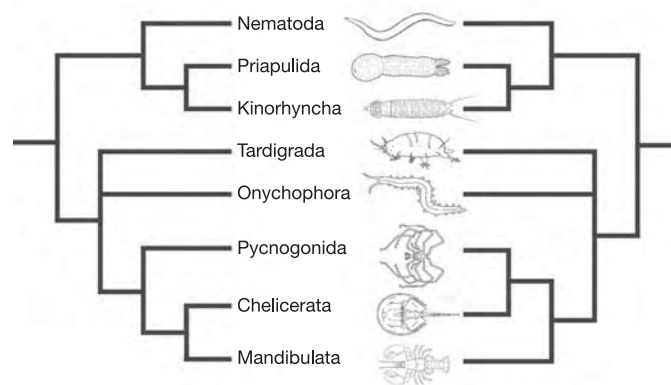


Figure 1 | Proposed relationships of pycnogonids among other ecdysozoans. The tree to the left reflects the Cormogonida hypothesis, with Pycnogonida as sister group to remaining extant arthropods. The tree to the right reflects the classical hypothesis of pycnogonids as sister group to chelicerates.

¹Department of Organismic & Evolutionary Biology and Museum of Comparative Zoology, Harvard University, 26 Oxford Street, ²Department of Organismic & Evolutionary Biology and Museum of Comparative Zoology, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138, USA. ³Kewalo Marine Lab, Pacific Biosciences Research Center, University of Hawaii, 41 Ahui Street, Honolulu, Hawaii 96813, USA.

larval appendage pair²². Of particular relevance with regard to the interpretation of CNS organization is that the body of the protonymphon effectively becomes the head of the adult²² and the protonymphon appendages correspond to the adult cephalic appendages (the chelifores, palps and ovigers; Fig. 2b)^{22,23}.

Immunostaining against acetylated α -tubulin revealed the axonal architecture of the developing protonymphon CNS, which is organized in a ring around the oesophagus (Fig. 3). Four bilateral pairs of distinct ganglia are connected by commissures across the oesophagus along the protonymphon anterior–posterior axis (Fig. 3a). The anteriormost supraoesophageal neuromere, the protocerebrum, consists of axon tracks that extend medially across the midline and receive the bifurcating nerves of the future optical apparatus dorsally, and of the cheliforal nerves anterolaterally (Fig. 3b, and Supplementary Movie 1). The anterior portion of each cheliforal ganglion innervates a chelifore by means of two nerve tracks (Fig. 3d, e), one innervating the fixed finger of the pincer, and the other the movable finger. In addition to the cheliforal ganglia at the lateral edges of the protocerebrum, the protonymphon CNS includes three posterior pairs of ganglia (Fig. 3a). The second (A2G) and third (A3G) pair are in close proximity and innervate the second and third appendages, respectively (Fig. 3c). The fibrous commissures between A2G appear to extend both above and below the oesophagus, whereas those of A3G extend only below. The terminal fourth pair of ganglia is positioned ventrolaterally at the base of the oesophagus (digestive tract is incomplete), at the posterior end of the protonymphon (Fig. 3d, f). Neuronal staining with anti-Elav (embryonic lethal abnormal visual system) was robust, and confirmed the above structural organization, including the ocular nerves at the dorsal surface of the protocerebrum (Supplementary Fig. 2).

Double immunostaining against acetylated α -tubulin and serotonin reveals a distinct subset of serotonergic neurons in the cheliforal ganglia (Fig. 3b), along two parallel tracks through the longitudinal connectives, and into the appendicular and posteriormost ganglia of the protonymphon CNS (Fig. 3f). In other arthropods examined, serotonergic neurons are identified most anteriorly in the median and lateral edges of the protocerebrum²⁴. Thus, anteriormost staining of serotonin in the cheliforal ganglia supports the association of these ganglia with the first neuromere of the arthropod CNS, the protocerebrum. Absence of reactivity in the medial region of the protocerebrum may be attributable to the delayed formation of the medial optical apparatus in the early protonymphon, in which the ocular nerves have only begun to develop above the protocerebral commissure.

Given the correspondence between the early protonymphon and the adult pycnogonid head²², and the resemblance in shape of the

protonymphon CNS to the larval brain ('neuropil ring') observed in other arthropods, we conclude that the three anterior pairs of ganglia in the early protonymphon CNS correspond to the tripartite, circumoral arthropod brain: the protocerebral commissure (presumably including the central/arcuate body), early ocular nerves and lateral cheliforal neuropils comprise the protocerebrum; the deutocerebrum, A2G; and the tritocerebrum, A3G. Because of the terminal position of the stomodeum, the longitudinal connectives encircling the oesophagus are oblique to the body axis (Supplementary Movies 1 and 2b, c) rather than bent dorsally in relation to the ventrally directed oesophagus as in other arthropods, such that the third poststomodeal commissure of the neuropil ring is suboesophageal rather than supraoesophageal (Fig. 3a, d). The proportionate size of the protonymphon neuromeres to one another is similar to those of other immature arthropods; the protocerebrum is significantly larger than both the deutocerebrum and the tritocerebrum⁸. The presence of strong reactivity to serotonin in the pair of suboesophageal posterior ganglia (Fig. 3f) indicates that the posteriormost protonymphon ganglia might correspond to the fourth cephalic ganglia, the mandibular ganglia of other arthropods²⁴.

The hypothesis that the pycnogonid protocerebrum consists of a single neuromere targeting the chelifores is the simplest interpretation of the data, requiring the fewest assumptions. However, owing to the minute size of the protonymphon, we are unable to eliminate alternative models unequivocally. Alternatives assume a hypothetical fusion of neuropils in the anteriormost region, such that either the protocerebrum, consisting only of the protocerebral commissure and ocular nerves, is fused to the cheliforal ganglia, which migrated from a deutocerebral position, or the protocerebral region includes two fused segmental neuromeres anterior to the deutocerebrum. In the first case, the assumptions are that the commissure between the ganglia targeting functional chelifores has not formed or is indistinguishable from the protocerebral commissure. Additionally, A2G must be interpreted as the tritocerebrum and thus the tritocerebral commissure as partly preoral, a condition not typical of arthropods. Conversely, if our initial interpretation is correct (A2G corresponds to the second neuromere) the commissure between A2G concurs with the proposal based on recent neuroembryological studies in Chelicerata¹ and Hexapoda²⁵ showing that the deutocerebral ganglia are transversely connected by preoral commissures as well as postoral fibres. The second alternative, in which the protocerebrum consists of two neuromeres, would support a controversial⁹ hypothesis based on the expression of segment polarity and dorsoventral patterning genes in insects claiming a bisegmental protocerebrum (labral and ocular neuromere)⁸. However, the arthropod protocerebrum generally includes a collection of neuropils that vary between taxa¹⁴ (for

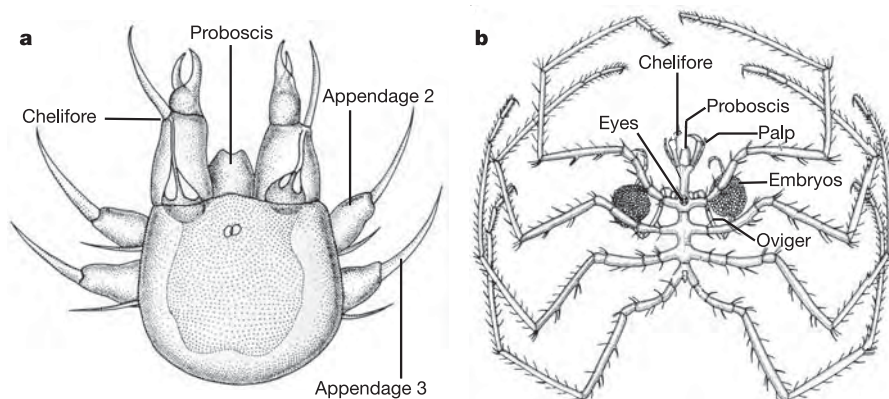


Figure 2 | The unique adult and larval morphology of pycnogonids. **a**, The three appendages of the protonymphon larva (shown) correspond to the cephalic appendages of the adult pycnogonid. (Modified from ref. 26.) **b**, The

adult male cares for embryos until hatching (*Nymphon rubrum*; modified from ref. 32).

example ‘ellipsoid body’, ‘paired noduli’), and there is no indication here that the ocular nerves and cheliforal ganglia belong to separate neuromeres.

Pycnogonids are often placed with chelicerates on the basis of the morphological similarities of their uniramous chelate (claw-like) appendages^{15,18}, although separate terms—pycnogonid chelifores versus chelicerate chelicerae—were explicitly chosen to reflect

uncertain homology^{23,26}. Positional support for homology has been based on classical descriptions of tritocerebral innervation of both chelifores and chelicerae in pycnogonids and arachnids^{15,27,28}. Recently, the chelicerae–tritocerebrum correspondence was disputed after the identification of deutocerebral ganglia targeting the chelicerae of horseshoe crabs¹. Evidence for tritocerebral innervation of the pycnogonid chelifores was presented in a descriptive model

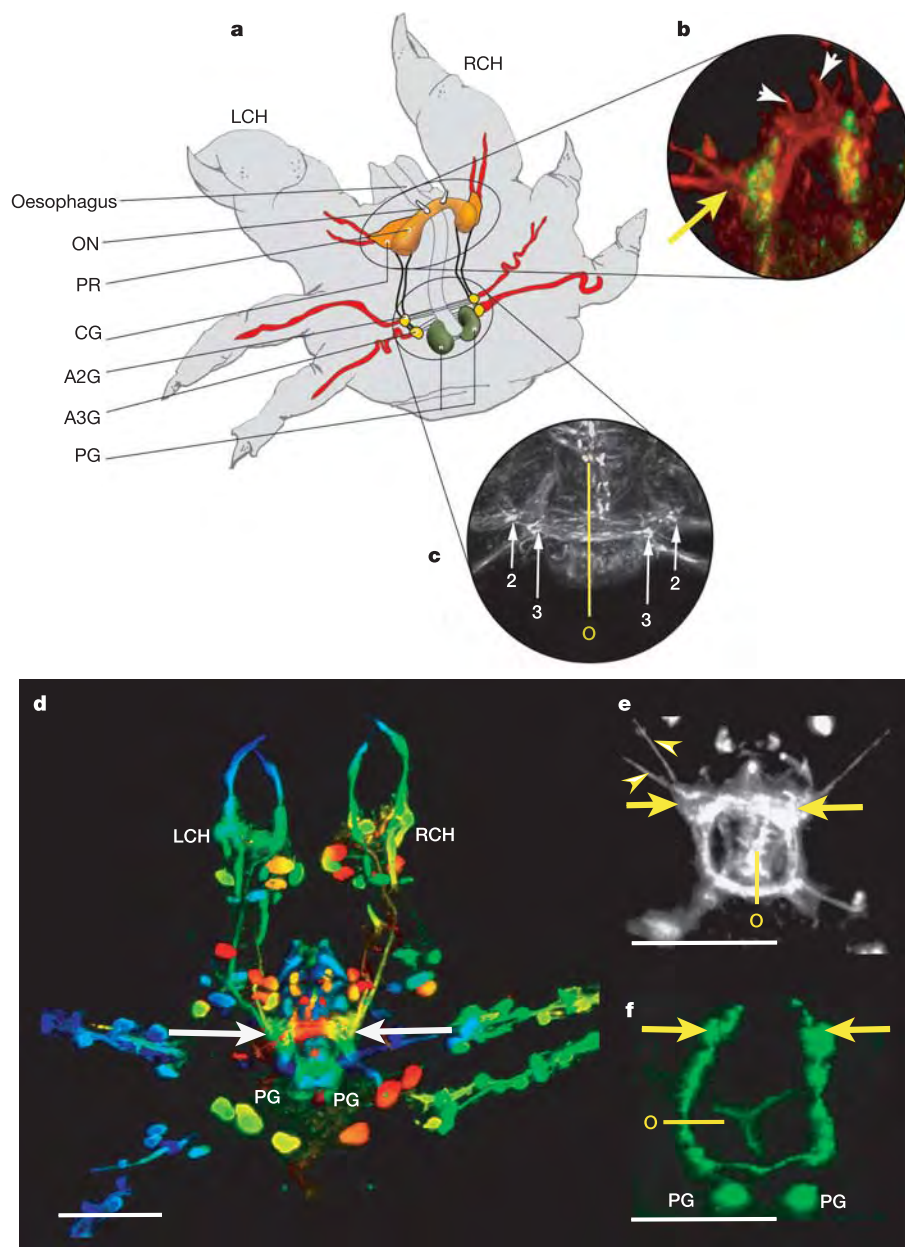


Figure 3 | CNS of the *Anoplodactylus* sp. larva (protonymph). **a**, Diagram of the protonymph seen from an oblique posteriolateral view based on reconstructions from confocal stacks (**b–f**). The CNS consists of four pairs of ganglia connected by commissures across the midline. The oesophagus runs through the proboscis between the left (LCH) and right (RCH) chelifore, and ends incompletely at the posterior ganglia (PG). The first neuromere is the protocerebrum, consisting of anteriolateral cheliforal ganglia (CG) connected by a prominent supraoesophageal protocerebral commissure (PR), and ocular nerves (ON). **b**, High magnification of the protocerebrum showing the cheliforal ganglion (arrows) and bifurcating ocular nerves (**b**, arrowheads). Circumoesophageal connectives run posteriorly from the PR, leading to the second (A2G) and third (A3G) ganglia that innervate the second and third

appendages (**c**). **c**, High magnification (α -tubulin, grey scale) of A2G and A3G and the fibrous appendicular commissures connecting the ganglia across the innervated oesophagus (O). **d**, Depth-coded image (α -tubulin): colours range from warm (red) indicating dorsal, to cooler (blue) indicating ventral. **e**, Transverse optical section (α -tubulin, grey scale) showing the circumoral shape of the protonymph ‘neuropil ring’ in relation to the oesophagus (O); the anterior bifurcating cheliforal nerves (arrowheads) target the cheliforal ganglia at the top of the ring. **f**, Same view of the CNS as in **e**, stained for serotonin (green). Immunoreactivity is visible in the cheliforal ganglia (arrows), along the circumoesophageal connectives, the suboesophageal appendicular commissure, and separately in the posterior ganglia (PG). Note background staining in the tripartite luminal surface of the oesophagus (O). Scale bars, 25 μ m.

by Wiren²⁸, and later affirmed by histology²⁷, in which a sub-pharyngeal commissure connecting the larval cheliforal ganglia (chelicerenneuromere) was illustrated as support for a postoral origin of the anterior chelifore ganglia. However, in his dissertation on pycnogonid development, T. H. Morgan²⁹ noted a discrepancy between innervation of chelifores and chelicerae: as embryos, arachnid chelicerae were innervated by postoral ganglia whereas pycnogonid chelifores arose heterolateral to the stomodeum and were not innervated by postoral ganglia. Our results confirm the observation that the commissure between the cheliforal ganglia is not postoral. The first set of midline commissures extending below the oesophagus are located between the second pair of ganglia, which innervate the second larval appendages. The chelifores clearly arise on segmental precursors anterior to the second larval appendages, and correspondingly to the first appendages of other extant arthropods, which are innervated by the second neuromere (chelicerae of chelicerates and (first) antennae of mandibulates)¹. In terms of structural similarity, chelate appendages have evolved convergently in many other arthropod groups, such as crustaceans, and appear on a variety of other segments, such as scorpion pedipalps¹².

This is direct neuroanatomical evidence of an appendage-bearing protocerebrum in an extant arthropod and corroborates previous data from onychophoran innervation¹⁹ and analyses of neuronal elements in *Drosophila* head gap gene mutants, used to reject the term 'acron' in favour of 'ocular segment'^{7,9}. The prostomium of annelids is derived from nonsegmental first-quartet micromere derivatives¹²; our data indicating a segmental origin of the anteriormost region in arthropods therefore further weaken the Articulata hypothesis.

In certain Palaeozoic stem-group arthropods, a pair of robust raptorial frontal appendages emerge from the anteriormost ocular region, theoretically innervated by the protocerebrum⁶. These frontal 'great appendages', are controversially thought to have been transformed into the labrum, a lip-like structure derived from the anterior portion of the arthropod head⁶. An argument has been revived for an association of the labrum with the protocerebral region based on the expression pattern of segment polarity genes during embryonic *Drosophila* development⁸. Pycnogonids do not possess a labrum, yet the homology of the labrum with pycnogonid chelifores is currently difficult to assess because of clashing views on labrum-neuromere innervation and the appendicular nature of the labrum. Further investigation with segment-specific candidate gene expression data might help to explain the potential homology of the labrum in pycnogonids. Apart from the labrum, the only possible segmental homologue of the chelifore is the frontal appendage found in stem-group arthropods such as *Kerygmachela*, *Leanchoilia* and basal anomalocaridids.

Innervation of the first appendages from the anteriormost part of the brain in onychophorans¹³ (sister group to arthropods), possibly in stem-group arthropods⁶, and in pycnogonids will affect future phylogenetic analyses by increasing the support for an independent clade of pycnogonids at the base of the arthropod tree, making the presence of a protocerebral appendage a plesiomorphic state. If pycnogonids are found to be sister group to chelicerates, and the protocerebral innervation of frontal appendages an ancestral feature, at least two independent transformations (if chelifores correspond to the labrum) or losses of the protocerebral appendages must be inferred within arthropod phylogeny. Alternatively, the possibility remains that chelifores are a pycnogonid apomorphy, and are thus a derived condition in arthropods that still supports previous suggestions that the protocerebrum maintains the potential to act as an appendage-bearing neuromere^{10,14}.

By investigating the CNS of a basal arthropod we have gained insight into the evolution of arthropod head segmentation and appendage correspondences. Pycnogonid chelifores and chelicerate chelicerae are convergent structures, innervated from different segmental neuromeres. Our finding supports previous analyses

predicting that arthropods once had 'great appendages' innervated by the protocerebrum, which have been lost or unrecognizably transformed in all extant arthropods except pycnogonids. This study further closes the gap in diagnosing positional homology between stem and crown-group arthropod head segments. Still, the defining criterion for homology, common ancestry, cannot be certain without a well-resolved phylogeny. If evidence is found sufficient to deem pycnogonids the most basal group of extant arthropods, our results support previous models of head evolution that predict that the original arthropod bore an acronless^{6,7,9,13} four-segmented head³⁰, encapsulating a tripartite circumoral brain rotated in an axial position, reminiscent of that found in onychophorans, nematodes and other cycloneurians^{13,19}.

METHODS

Specimen collection. Protonymphs of *Anoplodactylus* sp. were reared from embryos collected from males living in Kewalo Basin, Oahu, Hawaii. Detailed examination of adults indicates that our specimens might belong to a new, as yet undescribed species (C. P. Arango, personal communication).

Immunohistochemistry of CNS. Neuronal subsets were labelled as described previously³¹, with the following modifications. Before labelling, specimens were treated with a proteinase K solution (10 µg ml⁻¹) for 5 min and bath-sonicated for 5 s at a low setting. Staining of microtubules, such as the axonema of neurons, was obtained with a monoclonal anti-acetylated α -tubulin antibody and, independently, with Elav (for each, mouse IgG; Developmental Studies Hybridoma Bank) at dilutions of 1:5 and 1:200, respectively, in block solution (PBS containing 0.1% Triton X-100 and 5% normal goat serum) overnight at 4 °C. Serotonergic neurons were labelled with monoclonal anti-serotonin-specific antibodies (rabbit IgG; Immunostar) diluted 1:100 in block solution, overnight at 4 °C. Specimens were incubated for 5 h at 27 °C in secondary antibodies, Cy3-conjugated goat anti-mouse (Jackson ImmunoResearch Laboratories) and Alexa-488-conjugated donkey anti-rabbit (Molecular Probes), at a dilution of 1:250. Omission of the primary antibody abolished staining. Images were taken with a Zeiss LSM510 META confocal laser-scanning microscope.

Received 10 February; accepted 30 June 2005.

- Mittmann, B. & Scholtz, G. Development of the nervous system in the 'head' of *Limulus polyphemus* (Chelicerata: Xiphosura): morphological evidence for a correspondence between the segments of the chelicerae and of the (first) antennae of Mandibulata. *Dev. Genes Evol.* **213**, 9–17 (2003).
- Goodrich, E. S. On the relation of the arthropod head to the annelid prostomium. *Q. J. Microsc. Sci.* **247**, 248–268 (1897).
- Telford, M. J. & Thomas, R. H. Expression of homeobox genes shows chelicerate arthropods retain their deutocerebral segment. *Proc. Natl Acad. Sci. USA* **95**, 10671–10675 (1998).
- Damen, W. G. M., Hausdorf, M., Seyfarth, E.-A. & Tautz, D. A conserved mode of head segmentation in arthropods revealed by the expression pattern of Hox genes in a spider. *Proc. Natl Acad. Sci. USA* **95**, 10665–10675 (1998).
- Whittington, P. M. in *Evolutionary Developmental Biology of Crustacea* (ed. Scholtz, G.) 135–167 (A. A. Balkema, Berlin, 2004).
- Budd, G. E. A paleontological solution to the arthropod head problem. *Nature* **417**, 271–275 (2002).
- Schmidt-Ott, U., Gonzalez-Gaitan, M. & Technau, G. M. Analysis of neural elements in head-mutant *Drosophila* embryos suggests a segmental origin of the optic lobes. *Wilhelm Roux Arch. Dev. Biol.* **205**, 31–44 (1995).
- Urbach, R. & Technau, G. M. Early steps in building the insect brain: neuroblast formation and segmental patterning in the developing brain of different insect species. *Arthropod Struct. Dev.* **32**, 103–123 (2003).
- Rogers, B. T. & Kaufman, T. C. Structure of the insect head as revealed by the EN protein pattern in developing embryos. *Development* **122**, 3419–3432 (1996).
- Boyan, G. & Williams, L. A single cell analysis of *engrailed* expression in the early embryonic brain of the grasshopper *Schistocerca gregaria*: ontogeny and identity of the secondary headspot cells. *Arthropod Struct. Dev.* **30**, 207–218 (2002).
- Giribet, G., Edgecombe, G. D. & Wheeler, W. Arthropod phylogeny based on eight molecular loci and morphology. *Nature* **413**, 157–161 (2001).
- Brusca, R. C. & Brusca, G. J. *Invertebrates* 2nd edn (Sinauer, Sunderland, Massachusetts, 2003).
- Eriksson, B. J., Tait, N. N. & Budd, G. E. Head development in the onychophoran *Euperipatoides kanagrensis* with particular reference to the central nervous system. *J. Morphol.* **255**, 1–23 (2003).
- Bullock, T. H. & Horridge, G. A. (eds) *Structure and Function in the Nervous Systems of Invertebrates* (W. H. Freeman & Company, San Francisco, 1965).

15. Dunlop, J. A. & Arango, C. P. Pycnogonid affinities: a review. *J. Zool. Syst. Evol. Res.* **43**, 8–21 (2005).
16. Regier, J. C., Shultz, J. W. & Kambic, R. E. Pancrustacean phylogeny: hexapods are terrestrial crustaceans and maxillopods are not monophyletic. *Proc. R. Soc. Lond. B* **272**, 395–401 (2005).
17. Waloszek, D. & Dunlop, J. A. A larval sea spider (Arthropoda: Pycnogonida) from the Upper Cambrian 'Orsten' of Sweden, and the phylogenetic position of pycnogonids. *Palaeontology* **45**, 421–446 (2002).
18. Siveter, D. J., Sutton, M. D., Briggs, D. E. G. & Siveter, D. J. A Silurian sea spider. *Nature* **431**, 978–980 (2004).
19. Eriksson, B. J. & Budd, G. E. Onychophoran cephalic nerves and their bearing on our understanding of head segmentation and stem-group evolution of Arthropoda. *Arthropod Struct. Dev.* **29**, 197–209 (2000).
20. Miyazaki, K. On the shape of the foregut lumen in sea spiders (Arthropoda: Pycnogonida). *J. Mar. Biol. Assoc. UK* **82**, 1037–1038 (2002).
21. Bain, B. A. Larval types and a summary of postembryonic development within the pycnogonids. *Invertebr. Reprod. Dev.* **43**, 193–222 (2003).
22. Vilpoux, K. & Waloszek, D. Larval development and morphogenesis of the sea spider *Pycnogonum litorale* (Ström, 1972) and the tagmosis of the body of Pantopoda. *Arthropod Struct. Dev.* **32**, 349–383 (2003).
23. King, P. E. *Pycnogonids* (Hutchinson & Co. Ltd., London, 1973).
24. Sandeman, D. C., Sandeman, R. E. & Aitken, A. R. Atlas of serotonin-containing neurons in the optic lobes and brain of the crayfish, *Cherax destructor*. *J. Comp. Neurol.* **269**, 465–478 (1988).
25. Boyan, G., Reichert, H. & Hirth, F. Commissure formation in the embryonic insect brain. *Arthropod Struct. Dev.* **32**, 61–77 (2003).
26. Hedgepeth, J. W. On the evolutionary significance of Pycnogonida. *Smithson. Misc. Coll.* **106**, 1–53 (1947).
27. Winter, G. Beiträge zur Morphologie und Embryologie des vorderen Körperabschnitts (Cephalosoma) der Pantopoda Gerstaecker, 1863. I. Entstehung und Struktur des Zentralnervensystems. *Z. Zool. Syst. Evolforsch.* **18**, 27–61 (1980).
28. Wirén, E. in *Zologiska Bidrag Från Uppsala* Vol. 6, 41–181 (Univ. Uppsala, Uppsala, 1918).
29. Morgan, T. H. A contribution to the embryology and phylogeny of the pycnogonids. *Stud. Biol. Johns Hopkins Univ.* **5**, 1–76 (1891).
30. Walossek, D. & Müller, K. J. Upper Cambrian stem-lineage crustaceans and their bearing upon the monophyly of Crustacea and the position of *Agnostus*. *Lethaia* **23**, 409–427 (1990).
31. Dickinson, A. J., Croll, R. P. & Voronezhskaya, E. E. Development of embryonic cells containing serotonin, catecholamines, and FMRFamide-related peptides in *Aplysia californica*. *Biol. Bull.* **199**, 305–315 (2000).
32. Fage, in *Traité de Zoologie: Anatomie-Systematique Biologie* (ed. Grasse, P. P.) 906–941 (Masson et Cie, Paris, 1949).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank W. Morrissey for diagram preparation; E. C. Seaver for assistance with immunohistochemistry; and J. Hanken, G. Das, G. Edgecombe and A. Hejnol for advice and discussion. We thank the Developmental Studies Hybridoma Bank for the anti-tubulin and Elav antibodies. This material is based on work supported by the National Science Foundation AToL program to G.G. and M.Q.M.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.M. (amaxmen@oeb.harvard.edu).

Adaptive evolution of non-coding DNA in *Drosophila*

Peter Andolfatto¹

A large fraction of eukaryotic genomes consists of DNA that is not translated into protein sequence, and little is known about its functional significance. Here I show that several classes of non-coding DNA in *Drosophila* are evolving considerably slower than synonymous sites, and yet show an excess of between-species divergence relative to polymorphism when compared with synonymous sites. The former is a hallmark of selective constraint, but the latter is a signature of adaptive evolution, resembling general patterns of protein evolution in *Drosophila*^{1,2}. I estimate that about 40–70% of nucleotides in intergenic regions, untranslated portions of mature mRNAs (UTRs) and most intronic DNA are evolutionarily constrained relative to synonymous sites. However, I also use an extension to the McDonald–Kreitman test³ to show that a substantial fraction of the nucleotide divergence in these regions was driven to fixation by positive selection (about 20% for most intronic and intergenic DNA, and 60% for UTRs). On the basis of these observations, I suggest that a large fraction of the non-translated genome is functionally important and subject to both purifying selection and adaptive evolution. These results imply that, although positive selection is clearly an important facet of protein evolution, adaptive changes to non-coding DNA might have been considerably more common in the evolution of *D. melanogaster*.

The high degree of protein sequence similarity between phenotypically diverged species has led some to propose that regulatory evolution may be of considerably more importance than protein evolution^{4,5}. Although most of the typical eukaryotic genome is comprised of non-coding DNA, comparatively little is known about the evolutionary forces acting on it. Some unknown fraction of the non-translated genome is presumed to be crucial for the regulation of gene expression. Most of our direct knowledge regarding the evolution of regulatory elements comes from a handful of direct functional studies^{5,6}. A second, indirect approach is based on comparative genomics⁷. The rationale for this second approach is that if newly arising mutations are typically detrimental to gene function, functionally important parts of the genome are expected to evolve more slowly than those lacking function^{8–11}.

There are some limitations to the comparative genomics approach. First, a given genomic region might be conserved owing simply to a lower mutation rate¹². Second, known regulatory elements do not seem to be particularly well conserved as a class, at least in *Drosophila*¹⁰. This finding suggests that taking an approach based on sequence conservation alone may lead to a biased view of regulatory evolution. Functionality of DNA sequences implies that they can be subject to both negative and positive selection. If a significant fraction of divergence between species observed in non-coding DNA is positively selected rather than selectively neutral or constrained, this could lead to underestimates of the functional importance of non-coding DNA and cause researchers to overlook the contribution of arguably the most interesting class of mutations in genome evolution—those reflecting adaptive differences between populations and species.

These limitations can be overcome by combining comparative genomic analyses with population-level variability data^{1–3,13}. To assess the mode of selection acting on non-coding DNA, I have analysed new and previously published polymorphism data for 35 coding fragments (average length 667 base pairs (bp)) and 153 non-coding fragments (average length 426 bp) scattered across the X chromosome of *D. melanogaster* (see Supplementary Materials 1). To estimate levels of between-species divergence, I have compared *D. melanogaster* with its closely related sibling species, *D. simulans*.

On the basis of the current *Drosophila* genome annotation (release 4), I separated the surveyed fragments into several categories that are likely to differ in the intensity and mode of selection acting on them (see Table 1). It is apparent that most non-coding DNA evolves considerably slower than synonymous sites (that is, sites in protein-coding sequences at which mutations do not result in amino acid substitutions; Table 1). This is the case for introns and UTRs (see also refs 14–16), as well as intergenic DNA, much of which is far from the closest known gene (see Supplementary Materials 1). I estimate levels of constraint in *Drosophila* non-coding DNA to be 40% for introns, 50% for intergenic regions (IGRs), and 60% for UTRs (Table 2). These are all considerably higher than previous estimates from a variety of species comparisons^{11,15–18}. The non-coding DNA surveyed is also generally less polymorphic than synonymous sites in *D. melanogaster* (Table 1; $p < 10^{-10}$, Wilcoxon two-sample test for UTRs and introns+IGRs versus synonymous sites). Thus, both polymorphism and divergence in non-coding DNA are significantly reduced relative to synonymous sites in *D. melanogaster*.

Reduced levels of polymorphism and divergence in non-coding DNA resemble general patterns of protein evolution¹⁹ and suggest that non-coding DNA is either functionally constrained or is subject to a lower mutation rate than synonymous sites. One way to distinguish between these two models is to consider the distribution of polymorphism frequencies. Negative selection acting on polymorphic variants will keep them at lower frequencies in a population than expected if they were neutral²⁰. Consistent with this prediction, the distribution of polymorphism frequencies at both non-coding DNA and amino acid sites is skewed towards rare frequencies relative to synonymous polymorphisms (as indicated by a more negative Tajima's *D* value²⁰, Fig. 1). The distribution of Tajima's *D* values for non-synonymous sites among loci is negatively skewed relative to synonymous sites, suggesting that amino acid polymorphisms are subject to purifying selection (Fig. 1; $p = 0.002$, Wilcoxon two-sample test versus synonymous sites). Here I show that this same pattern extends to polymorphisms in non-coding DNA (Fig. 1; Wilcoxon test versus synonymous sites: pooled non-coding, $p = 0.0001$; UTRs, $p < 0.0001$; introns, $p = 0.001$; IGRs, $p = 0.005$). This finding, together with the observed reduction in polymorphism and divergence, implies that mutations in non-coding DNA are subject, on average, to stronger negative selection than synonymous sites (see also Supplementary Materials 2).

Does selective constraint alone account for patterns of non-coding DNA evolution? McDonald and Kreitman³ have proposed a frame-

¹Section of Ecology, Behavior and Evolution, Division of Biological Sciences, University of California San Diego, La Jolla, California 92093, USA.

Table 1 | Polymorphism and divergence in coding and non-coding DNA of *D. melanogaster*

Mutation class	No. of regions	Mean π^*	Mean $D_{xy}^\dagger$	$D^\ddagger$	PS	$p $	$P^\P$	$p^\#$
Synonymous	35	2.87	13.59	604	502	—	323	—
Non-synonymous	35	0.18	1.72	260	115	$<10^{-6}$	52	$<10^{-9}$
Non-coding	153	1.06	5.94	3,168	2,386	0.14	1,295	$<10^{-3}$
UTRs	31	0.54	4.54	471	246	$<10^{-5}$	107	$<10^{-11}$
5'UTRs	18	0.61	5.41	328	160	$<10^{-5}$	71	$<10^{-9}$
3'UTRs	13	0.45	3.35	143	86	0.034	36	$<10^{-4}$
Introns	72	1.25	6.71	1,564	1,221	0.39	675	0.010
IGRs	50	1.11	5.72	1,133	919	>0.5	513	0.059
pIGRs	20	1.29	6.58	500	400	>0.5	237	0.25
dIGRs	30	0.99	5.18	633	519	>0.5	276	0.041
Introns+IGR	122	1.19	6.25	2,697	2,140	0.50	1,188	0.013

Mutation classes: synonymous sites, non-synonymous sites, untranslated transcribed regions (UTRs), intergenic regions within 2 kb of a gene (pIGRs), intergenic regions more than 4 kb away from a gene (dIGRs).
* π is the weighted average within-species pairwise diversity per 100 sites.
 $^\dagger D_{xy}$ is the weighted average pairwise divergence per 100 sites between *D. melanogaster* and *D. simulans*, corrected for multiple hits (Jukes-Cantor). D_{xy} at fourfold degenerate synonymous sites is 12.0%.
 $^\ddagger D$ is the estimated number of fixed differences between species using a Jukes-Cantor correction for multiple hits (see Methods).
 $^\P P$ is the number of intraspecific polymorphisms.
 $||$ McDonald-Kreitman test of probability using all polymorphisms.
 $^\P P$ is the number of intraspecific polymorphisms excluding singletons.
 $^\#$ McDonald-Kreitman test of probability excluding singleton polymorphisms. Probabilities are from two-tailed Fisher's exact tests and assume sites are independent. These are likely to be only slight underestimates given probable levels of intragenic recombination (see Supplementary Materials 2).

work to distinguish neutrality (and variation in mutation rate) from negative and positive selection in the genome. Their approach compares levels of polymorphism within and divergence between species for a putatively selected class of sites in the genome to a neutral standard. If reduced levels of polymorphism and divergence in non-coding DNA can be explained by a lower mutation rate, the ratio of polymorphism to divergence should be similar to that for synonymous sites. Positive selection will increase divergence relative to polymorphism at selected sites, whereas negative selection is expected to result in the opposite pattern²¹. Although this framework was originally designed to detect selection within protein-coding genes, it can be generalized to consider arbitrary classes of putatively selected sites sampled from multiple genomic regions, including non-coding DNA (see Supplementary Materials 2). Using all polymorphisms, there is a significant excess of divergence for amino acid replacement sites ($p = 5 \times 10^{-7}$) and for UTRs ($p = 3 \times 10^{-6}$, two-tailed Fisher's exact test) but not at other subclasses of non-coding DNA (Table 1). This preliminary analysis suggests that, similar to the pattern observed for amino acid substitutions^{1,2}, a significant proportion of nucleotide divergence at UTRs was also driven to fixation by positive selection.

The presence of weakly negatively selected variants in polymorphism can mask the signature of adaptive evolution in the genome^{1,22}, making the McDonald-Kreitman test very conservative. As I have shown above that polymorphic variants in non-coding DNA are subject to stronger selective constraint than synonymous sites (Table 1 and Fig. 1), negatively selected variants contributing to polymorphism in non-coding DNA are likely to be a factor limiting

power to detect positive selection. This problem can be partially overcome by considering only those mutations that are not rare in a sample from both the neutral and putatively selected classes (see ref. 23 and Supplementary Materials 2). Applying this approach reveals a significant excess of divergence in UTRs and in most other classes of non-coding DNA relative to synonymous sites (Table 1; UTRs, $p = 5 \times 10^{-12}$; introns, $p = 0.01$; dIGRs, $p = 0.04$; introns+IGRs, $p = 0.01$). A Hudson-Kreitman-Aguadé (HKA) test²⁴ also provides statistical support for a reduced ratio of polymorphism to divergence for non-coding DNA relative to synonymous sites (UTRs, $p < 10^{-3}$; pooled introns and IGRs, $p = 0.02$; see Supplementary Materials 2). Together, these results show that a significant fraction of the divergence in UTRs, introns and intergenic DNA was probably driven to fixation by positive selection.

To quantify the intensity and the relative importance of positive selection in shaping the evolution of non-coding DNA, I apply two extensions of the McDonald-Kreitman approach^{2,13}. First I estimate α , defined as the proportion of the divergence between species that was driven by positive selection². I estimate that about 20% of the nucleotide divergence in introns and intergenic DNA was driven to fixation by positive selection, and about 60% for UTRs (Fig. 2a and Table 2). Using a hierarchical bayesian framework¹³, I estimate the

Table 2 | Functionally relevant nucleotides in non-coding DNA

Class	C (%) [*]	α (%) [†]	$p(\alpha \leq 0)$ [‡]	FRN (%) [§]
UTRs	60.4	57.5	$<10^{-3}$	83.2
5'UTRs	52.9	60.8	$<10^{-3}$	80.9
3'UTRs	70.7	52.9	$<10^{-3}$	86.2
Introns	39.5	19.3	0.007	51.2
IGRs	49.3	15.3	0.036	57.1
pIGRs	40.6	11.4	0.165	47.4
dIGRs	54.6	18.5	0.019	63.0
Introns + IGR	44.2	17.6	0.013	54.0

*Constraint (C) is estimated relative to fourfold degenerate synonymous sites.
 $^\dagger \alpha$ is the estimated fraction of divergence driven by positive selection.
 ‡ Probabilities ($\alpha \leq 0$) have been adjusted for effects of linkage within loci (see Supplementary Materials 2.5).
 § FRN is the inferred fraction of functionally relevant nucleotides given levels of constraint and α (that is, $FRN = C + (1 - C)\alpha$).

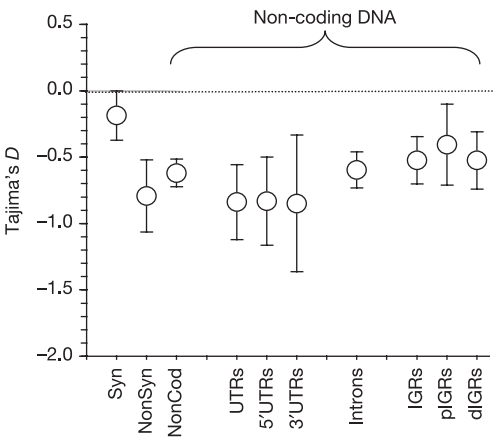


Figure 1 | Mean Tajima's D values for coding and non-coding DNA. Means across loci are given with bars indicating two standard errors. The expectation of D under the neutral model is shown as a dotted line. Syn, synonymous sites; NonSyn, non-synonymous sites; NonCod, pooled non-coding DNA.

selection intensity on non-coding DNA (including UTRs, introns and IGRs) to be positive and significantly different from zero in most cases (Fig. 2b; Supplementary Materials 3). As this bayesian approach assumes that segregating and fixed variants are subject to the same direction and intensity of selection, it is likely to underestimate the magnitude of $2N_e s$ (the intensity of selection) for nucleotide substitutions fixed by positive selection (see Supplementary Materials 2).

Evidence that a significant fraction of non-coding DNA is functionally important is emerging from a variety of comparative genomic studies. However, my finding of a large fraction of positively selected divergence implies that 'evolutionary constraint' will substantially underestimate the fraction of functionally relevant nucleotides because it ignores the contribution of positively selected mutations to divergence. For the example of UTRs, I estimate evolutionary constraint to be 60%. However, as 58% of the observed divergence was positively selected, this implies that 83% of nucleotides in UTRs are in fact functionally relevant. Likewise, the fraction of functionally relevant nucleotides in introns and IGRs is likely to be about 10–20% higher than suggested by levels of constraint alone (Table 2).

How frequent is adaptation in the *Drosophila* genome? Rough calculations (see Supplementary Materials 4) suggest that there has been about one adaptive amino acid substitution every 20 years since the split of *D. melanogaster* and *D. simulans* (see also ref. 2). Although this is substantial, consider that the total number of sites contained in

introns, intergenic regions and UTRs far outweighs the number of codons in the *Drosophila* genome²⁵. I estimate that UTRs alone contribute as much to adaptive divergence between species as do amino acid changes, and the summed contribution of non-coding DNA to adaptive divergence could easily be an order of magnitude larger. These findings support previous intuitions^{4,5} about the great importance of regulatory changes in evolution.

METHODS

Data. All loci used in this study, previously published or newly collected, are X-linked genomic fragments, with a sample size of 12 *D. melanogaster* alleles sampled from a population in Zimbabwe, and a single *D. simulans* sequence. For coding DNA (synonymous and non-synonymous sites), I collected polymorphism and divergence in 31 coding regions selected randomly with respect to gene function, and 51 non-coding regions (27 intergenic and 24 untranslated transcribed regions). Information about these 82 loci and primers used can be found in Supplementary Materials 1. I used polymerase chain reaction (PCR) to amplify 700–800-bp regions from genomic DNA extracted from single male flies, removed primers and nucleotides using exonuclease I and shrimp alkaline phosphatase, and sequenced the cleaned product on both strands using Big-Dye (Version 3, Applied Biosystems). Sequences were collected on an ABI 3730 capillary sequencer and were aligned and edited using the program Sequencher (Gene Codes).

To the 82 regions surveyed above, I added previously published data for loci that had the same sample size ($n = 12$ flies) and were surveyed in similar samples from Zimbabwe^{26,27}. A number of the previously published loci²⁶ had to be functionally reassigned when compared to Release 4 of the annotated *D. melanogaster* genome (<http://flybase.bio.indiana.edu/annot/dmel-release4.html>). I excluded any loci in regions of reduced recombination (see below). Previously published loci fitting these requirements were processed into 106 fragments (4 coding, 7 UTR, 23 intergenic and 72 intron). Thus, the total number of regions surveyed in this analysis is 188. Alignments for each locus are available upon request. A reciprocal best-hit BLAST protocol was used to confirm that the regions compared between *D. melanogaster* and *D. simulans* are indeed orthologous. Extra gaps were introduced into some alignments in regions that were particularly difficult to align. This procedure is likely to upwardly bias estimates of constraint, but is conservative with respect to detecting positive selection.

Analyses. The estimated number of synonymous sites, non-synonymous sites, average pairwise diversity (π), average pairwise divergence (D_{xy}), as well as counts of the number of polymorphic sites (P) were performed using DnaSP software (version 4; <http://www.ub.es/dnasp/>) and Perl code written by P.A. The number of divergent sites (D) was estimated as $D_{xy} - \pi$ using a Jukes–Cantor correction for multiple hits. Multiply hit sites were included in the analysis but insertion–deletion polymorphisms and mutations overlapping alignment gaps were excluded. Derived mutations were polarized using a single *D. simulans* sequence and assuming standard parsimony criteria. Tajima's D value²⁰ was estimated from the number of polymorphisms and π .

In this study, I assume that synonymous sites are more neutral than putatively selected classes of sites (see Supplementary Materials 2.2). I separated non-coding DNA into subclasses that I expected *a priori* to experience different selection pressures: 5' and 3' untranslated transcribed regions (UTRs), introns, intergenic regions within 2 kilobases (kb) of a gene (proximal intergenic regions, pIGRs), and intergenic regions further than 4 kb from the nearest gene (distal intergenic regions, dIGRs). My sample of intron fragments is biased towards introns larger than the median intron size (86 bp) (ref. 28), making estimates of constraint higher than expected with a random sample of introns¹⁴. However, 95% of intronic DNA is contained within introns longer than the median size²⁸, and thus my estimate reflects levels of constraint for most intronic DNA in the genome.

For comparisons of polymorphism and divergence between synonymous sites and non-coding DNA, it was necessary to pool sites in each class. I estimate evolutionary constraint relative to fourfold degenerate synonymous sites using the approach in ref. 15, except that I pooled classes of sites and used a Jukes–Cantor correction for multiple hits¹⁹. Given differences in base composition between coding and non-coding regions, I investigated possible differences in mutations rates owing to the 16 possible adjacent-base contexts of nucleotides (suggested by A. Kondrashov). There was no significant effect of adjacent-base context on rates of divergence (see Supplementary Materials 5).

I estimate the proportion of divergence driven by positive selection^{1,2} as $\alpha = 1 - (D_S P_X / D_X P_S)$, where S denotes synonymous (that is, putatively neutral) sites, X denotes putatively selected sites, and $D = \sum_{i=1}^n D_i$ and $P = \sum_{i=1}^n P_i$,

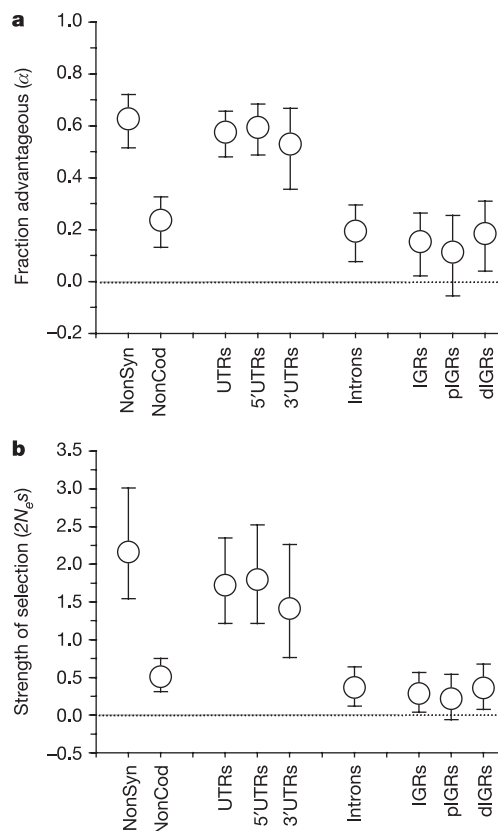


Figure 2 | Quantifying adaptive divergence and selection intensity.

a, Estimates of α , the fraction of nucleotide divergence driven by positive selection. Error bars indicate 90% confidence limits determined by a non-parametric bootstrapping. Estimated probabilities that $\alpha \geq 0$ corrected for partial linkage are given in Table 2. **b**, Estimates of the intensity of selection ($2N_e s$) acting on non-synonymous and non-coding DNA sites. Error bars indicate 90% confidence limits determined by simulation (see Methods). Singleton polymorphisms were excluded in estimates of α and $2N_e s$ (see Supplementary Materials 3). Abbreviations as in Fig. 1.

where D_i and P_i are the number of divergent and polymorphic variants at locus i , respectively, and n is the number of loci of class S or X. Confidence limits on α were estimated using a standard non-parametric bootstrapping procedure, assuming sites are independent. The issue of non-independence of sites within surveyed fragments is addressed in Supplementary Materials 2.5. For consistency, α was estimated for non-synonymous sites in the same way. The intensity of selection ($2N_e s$) was estimated on putatively selected classes (pooling sites as above) using a hierarchical bayesian method (<http://cbsuapps.tc.cornell.edu>)¹³. To avoid problems associated with large-scale variation in recombination rates, I restricted my survey of loci to regions of the X chromosome that have the highest rates of recombination²⁹ (see Supplementary Fig. 1.1).

Received 23 May; accepted 2 August 2005.

- Fay, J. C., Wyckoff, G. J. & Wu, C. I. Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. *Nature* **415**, 1024–1026 (2002).
- Smith, N. G. & Eyre-Walker, A. Adaptive protein evolution in *Drosophila*. *Nature* **415**, 1022–1024 (2002).
- McDonald, J. & Kreitman, M. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**, 652–654 (1991).
- King, M. C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).
- Carroll, S. B., Grenier, J. K. & Weatherbee, S. D. *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design* (Blackwell Science, Malden, Massachusetts, 2001).
- Ludwig, M. *et al.* Functional evolution of a *cis*-regulatory module. *PLoS Biol.* **3**, e93 (2005).
- Miller, W., Makova, K., Nekrutenko, A. & Hardison, R. Comparative genomics. *Annu. Rev. Genomics Hum. Genet.* **5**, 15–56 (2004).
- Cliften, P. *et al.* Surveying *Saccharomyces* genomes to identify functional elements by comparative DNA sequence analysis. *Genome Res.* **11**, 1175–1186 (2001).
- Gibbs, R. *et al.* Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**, 493–521 (2004).
- Richards, S. *et al.* Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and *cis*-element evolution. *Genome Res.* **15**, 1–18 (2005).
- Shabalina, S. & Kondrashov, A. Pattern of selective constraint in *C. elegans* and *C. briggsae* genomes. *Genet. Res.* **74**, 23–30 (1999).
- Clark, A. The search for meaning in noncoding DNA. *Genome Res.* **11**, 1319–1320 (2001).
- Bustamante, C. *et al.* The cost of inbreeding in *Arabidopsis*. *Nature* **416**, 531–534 (2002).
- Haddrill, P. R., Halligan, D., Charlesworth, B. & Andolfatto, P. Patterns of intron sequence evolution in *Drosophila* are dependent upon length and GC content. *Genome Biol.* **6**, R67 (2005).
- Halligan, D., Eyre-Walker, A., Andolfatto, P. & Keightley, P. Patterns of evolutionary constraints in intronic and intergenic DNA of *Drosophila*. *Genome Res.* **14**, 273–279 (2004).
- Bachtrog, D. Sex chromosome evolution: molecular aspects of Y chromosome degeneration in *Drosophila*. *Genome Res.* **15**, 1393–1401 (2005).
- Jareborg, N., Birney, E. & Durbin, R. Comparative analysis of noncoding regions of 77 orthologous mouse and human gene pairs. *Genome Res.* **9**, 815–824 (1999).
- Bergman, C. & Kreitman, M. Analysis of conserved noncoding DNA in *Drosophila* reveals similar constraints in intergenic and intronic sequences. *Genome Res.* **11**, 1335–1345 (2001).
- Li, W. *Molecular Evolution* (Sinauer Associates, Sunderland, Massachusetts, 1997).
- Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595 (1989).
- Kimura, M. *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, 1983).
- Charlesworth, B. The effect of background selection against deleterious mutations on weakly selected, linked variants. *Genet. Res.* **63**, 213–227 (1994).
- Templeton, A. Contingency tests of neutrality using intra/interspecific gene trees: the rejection of neutrality for the evolution of the mitochondrial cytochrome oxidase II gene in the hominoid primates. *Genetics* **144**, 1263–1270 (1996).
- Hudson, R., Kreitman, M. & Aguadé, M. A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–159 (1987).
- Misra, S. *et al.* Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review. *Genome Biol.* **3**, research0083.1-0083.22 (2002).
- Glinka, S., Ometto, L., Mousset, S., Stephan, W. & De Lorenzo, D. Demography and natural selection have shaped genetic variation in *Drosophila melanogaster*: a multi-locus approach. *Genetics* **165**, 1269–1278 (2003).
- Haddrill, P. R., Thornton, K. R., Charlesworth, B. & Andolfatto, P. Multilocus patterns on nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations. *Genome Res.* **15**, 790–799 (2005).
- Yu, J. *et al.* Minimal introns are not “junk”. *Genome Res.* **12**, 1185–1189 (2002).
- Charlesworth, B. Background selection and patterns of genetic diversity in *Drosophila melanogaster*. *Genet. Res.* **68**, 131–149 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The author thanks D. Bachtrog for extensive comments on the manuscript and help with data quality issues, C. Bustamante and K. Thornton for providing code, and B. Ballard for Zimbabwe fly lines. P. Haddrill and K. Thornton assisted in designing primers for distal intergenic and coding regions, respectively. Thanks to B. Fischman for technical help, A. Betancourt, A. Kondrashov, A. Poon, D. Presgraves, M. Przeworski and S. Wright for critical comments on the manuscript, and L. Chao and J. Huelsenbeck for advice. Thanks also to the Washington University Genome Sequencing Center for providing unpublished *D. simulans* sequences. This work was funded in part by a research grant from the Biotechnology and Biological Sciences Research Council (UK) to P.A. The author is supported by an Alfred P. Sloan Fellowship in Molecular and Computational Biology.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The author declares no competing financial interests. Correspondence and requests for materials should be addressed to P.A. (pandolfatto@ucsd.edu).

Natural selection on protein-coding genes in the human genome

Carlos D. Bustamante¹, Adi Fledel-Alon¹, Scott Williamson¹, Rasmus Nielsen^{1,2}, Melissa Todd Hubisz¹, Stephen Glanowski³, David M. Tanenbaum³, Thomas J. White⁴, John J. Sninsky⁴, Ryan D. Hernandez¹, Daniel Civello⁴, Mark D. Adams⁵, Michele Cargill^{4*} & Andrew G. Clark^{6*}

Comparisons of DNA polymorphism within species to divergence between species enables the discovery of molecular adaptation in evolutionarily constrained genes as well as the differentiation of weak from strong purifying selection^{1–4}. The extent to which weak negative and positive darwinian selection have driven the molecular evolution of different species varies greatly^{5–16}, with some species, such as *Drosophila melanogaster*, showing strong evidence of pervasive positive selection^{6–9}, and others, such as the selfing weed *Arabidopsis thaliana*, showing an excess of deleterious variation within local populations^{9,10}. Here we contrast patterns of coding sequence polymorphism identified by direct sequencing of 39 humans for over 11,000 genes to divergence between humans and chimpanzees, and find strong evidence that natural selection has shaped the recent molecular evolution of our species. Our analysis discovered 304 (9.0%) out of 3,377 potentially informative loci showing evidence of rapid amino acid evolution. Furthermore, 813 (13.5%) out of 6,033 potentially informative loci show a paucity of amino acid differences between humans and chimpanzees, indicating weak negative selection and/or balancing selection operating on mutations at these loci. We find that the distribution of negatively and positively selected genes varies greatly among biological processes and molecular functions, and that some classes, such as transcription factors, show an excess of rapidly evolving genes, whereas others, such as cytoskeletal proteins, show an excess of genes with extensive amino acid polymorphism within humans and yet little amino acid divergence between humans and chimpanzees.

Of considerable interest to medical geneticists is the extent to which weak negative selection has shaped the human genome. Sensitivity to weak selection may be important in identifying candidate genes for association mapping studies, because weakly deleterious mutations can reach appreciable frequencies in local populations and, thus, may contribute significantly to genetic variance in disease susceptibility. A team at Celera Genomics sequenced by exon-specific polymerase chain reaction (PCR) amplification 20,362 loci in 20 European Americans, 19 African Americans and one male chimpanzee with the initial intention of finding novel non-synonymous single nucleotide polymorphisms (SNPs) based on their 2001 build of the human genome. Here we analyse 11,624 genes with complementary DNA support, strong evidence of orthology between humans and chimpanzee, and location in non-repetitive elements of the genome. The distributions of the total number of synonymous and non-synonymous SNPs and fixed differences for these loci are presented in Fig. 1a. We found that 10,767 genes (92.6%) showed

some form of coding nucleotide variability either within human subjects or between humans and a chimpanzee. A total of 34,099 fixed synonymous differences between all humans in our sample and the chimpanzee yield a genomic average synonymous divergence of $\bar{d}_S = 1.02\%$. Correspondingly, we found 20,467 non-synonymous differences ($\bar{d}_N = 0.242\%$) across 11.81 megabases (Mb) of aligned coding DNA. We also discovered 15,750 synonymous and 14,311 non-synonymous SNPs among the human subjects, yielding average synonymous and non-synonymous SNP densities of $\bar{p}_S = 0.470\%$ and $\bar{p}_N = 0.169\%$. We note that the ratio of non-synonymous to synonymous differences (23.76%) is smaller than the ratio of non-synonymous to synonymous polymorphisms (38.42%), indicating a highly significant excess of amino acid variation relative to divergence ($\chi^2 = 816.03$, $P < 2 \times 10^{-16}$). This is consistent with previous studies of human polymorphism that suggest that a large proportion of amino acid variation in the human genome is slightly to moderately deleterious^{11–16}.

In order to identify particular genes with evidence of non-neutral evolution, we applied a statistical approach that makes efficient use of comparative population genomic data^{4,7,9,17}. The method quantifies the extent and directionality of selection operating on a given gene in terms of the population genetic selection parameter $\gamma = 2N_e s$ (where N_e is the effective population size and s is the selection coefficient in a Wright–Fisher genic selection model) as estimated from contingency tables comparing polymorphism (P) versus divergence (D) at synonymous (S) versus non-synonymous (N) sites^{3,4}. The parameter is negative if a gene shows an excess of amino acid polymorphism (or paucity of amino acid divergence) and positive if a gene has an excess of amino acid divergence relative to the genomic average for synonymous sites. Invariant sites in an alignment do not affect the estimate of γ , so it is independent of strong purifying selection operating at the locus. One consequence of this robustness is that genes with little or no variation contain little information regarding γ . We therefore concentrate on two subsets of the data: the $n = 3,277$ genes with at least four variable non-synonymous sites in the alignment (that is, $P_N + D_N \geq 4$), which are potentially informative about positive selection (IPS data), and the set with at least two variable non-synonymous sites, which are potentially informative only about negative selection (INS data; $n = 6,033$). In order to classify genes, we use the posterior distribution of the selection coefficient to calculate a test statistic, such that $P^+ = P_r\{\gamma > 0 | \text{Data}\}$ is the posterior probability that a gene is subject to positive darwinian selection and $P^- = P_r\{\gamma < 0 | \text{Data}\}$ is the posterior probability that a gene harbours excess amino acid variation. (Negative

¹Department of Biological Statistics and Computational Biology, 101 Biotechnology Building, Cornell University, Ithaca, New York 14853, USA. ²Center for Bioinformatics, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen, Denmark. ³Applied Biosystems, 45 West Gude Drive, Rockville, Maryland 20850, USA. ⁴Celera Diagnostics, 1401 Harbor Bay Parkway, Alameda, California 94502, USA. ⁵Department of Genetics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, Ohio 44106, USA. ⁶Department of Molecular Biology and Genetics, 227 Biotechnology Building, Cornell University, Ithaca, New York 14853, USA.

*These authors contributed equally to this work.

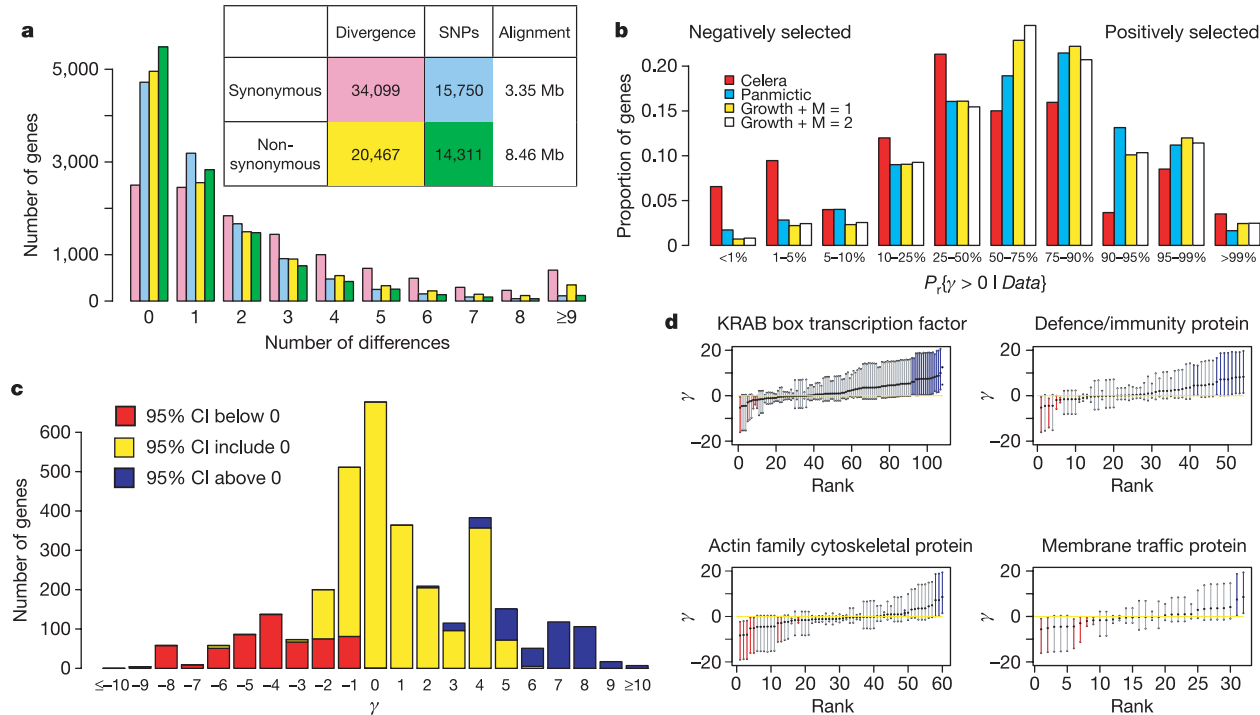


Figure 1 | Summary distributions of McDonald-Kreitman cell entries and mkprf analyses. **a**, Distributions of synonymous and non-synonymous SNPs and fixed differences across 11,624 genes. **b**, Distribution of the posterior probability that a gene is positively selected for simulated data under three neutral demographic scenarios and for the Celera data conditioning on a gene having at least four variable amino acid positions (IPS data). **c**, Distribution of the estimated average selection coefficient for the 3,277 genes in the Celera IPS data set. **d**, Distribution of γ for four molecular functions showing an excess of non-neutral loci. Bars represent 95% CIs with blue and red denoting CIs completely above or below $\gamma = 0$.

Table 1 | Molecular functions and biological processes showing as an excess of positively or negatively selected genes

Category	P-value	Number where CI > 0	Number where CI < 0	N
Biological process				
Apoptosis	0.00336	12	10 (5)	99 (53)
Cell structure and motility	0.00008	4	27 (8)	176 (101)
Ectoderm development	0.02805	1	12 (7)	98 (61)
Gametogenesis	0.03411	5	4 (1)	41 (23)
General vesicle transport	0.00016	0	14 (4)	40 (20)
Intracellular protein traffic	0.01151	8	32 (10)	159 (83)
mRNA transcription	0.00002	29	34 (17)	333 (185)
Natural-killer-cell-mediated immunity	0.03299	1	3 (2)	19 (9)
Nucleoside, nucleotide and nucleic acid metabolism	0.00467	38	65 (28)	568 (311)
Sensory perception	0.04577	9	9 (4)	101 (56)
Molecular function				
Actin family cytoskeletal protein	0.00008	4	23 (10)	104 (60)
Cytoskeletal protein	0.00000	7	36 (12)	205 (118)
Defence/immunity protein	0.00965	10	12 (5)	89 (54)
Extracellular matrix	0.01478	3	15 (11)	103 (78)
Homeotic transcription factor	0.02586	2	2 (0)	28 (8)
Immunoglobulin receptor family member	0.04558	7	3 (1)	36 (27)
Kinase modulator	0.00538	0	9 (3)	28 (14)
KRAB box transcription factor	0.00004	17	10 (5)	168 (108)
Membrane traffic protein	0.02701	2	17 (6)	70 (32)
Microtubule-binding motor protein	0.03109	1	7 (1)	31 (19)
Microtubule family cytoskeletal protein	0.01373	2	9 (1)	52 (34)
Non-motor actin-binding protein	0.01691	3	12 (3)	58 (27)
Nuclear hormone receptor	0.00143	3	0 (0)	10 (7)
Protein kinase	0.04366	6	4 (1)	94 (36)
Receptor	0.00211	31	37 (18)	343 (207)
RNA helicase	0.03948	0	8 (4)	33 (18)
Transcription factor	0.00000	39	41 (19)	428 (240)
Voltage-gated potassium channel	0.03943	0	5 (2)	23 (11)
Zinc finger transcription factor	0.00074	20	19 (9)	229 (141)

Classification is based on Panther classification. The P-value is the uncorrected value from the Mann-Whitney U-test. A total of 139 different molecular functions and 133 biological processes were tested; non-significant categories are listed in Supplementary Table 1, as are significant categories with considerable overlap with those shown in the Table. Parentheses for all count data denote the IPS data set. Bold text indicates negatively selected genes; non-bold text indicates positively selected genes.

selection on slightly deleterious mutations increases the proportion of non-synonymous polymorphisms relative to the proportion of non-synonymous fixed differences.)

Figure 1b, c shows histograms of the distribution of P^+ and γ among genes in the IPS data. In order to assess how deviations from the assumptions of our model may influence our conclusions, we simulated 10,000 replicate data sets under each of three different demographic models (See Supplementary Data 1 for details) and plotted the distribution of P^+ for the corresponding IPS data sets in these simulations. The Celera data have a clear excess of genes with high and low posterior probabilities (that is, there are too many genes

in the <1%, 1–5% and >99% categories) regardless of which demographic model is used as the null hypothesis. The signature is particularly strong for weak negative selection. Figure 1c shows the distribution of γ among genes in the IPS data set (estimated as the posterior mean of the selection parameter for each gene) and Fig. 1d shows the posterior mean and credibility intervals (CIs) for genes in four Panther classifications enriched for positively or negatively selected genes (see Table 1 and Supplementary Table 1 for details).

In Fig. 2 we present an amino acid selection map of the human genome with each locus mapped onto its genomic location and coloured according to the strength of evidence for positive or

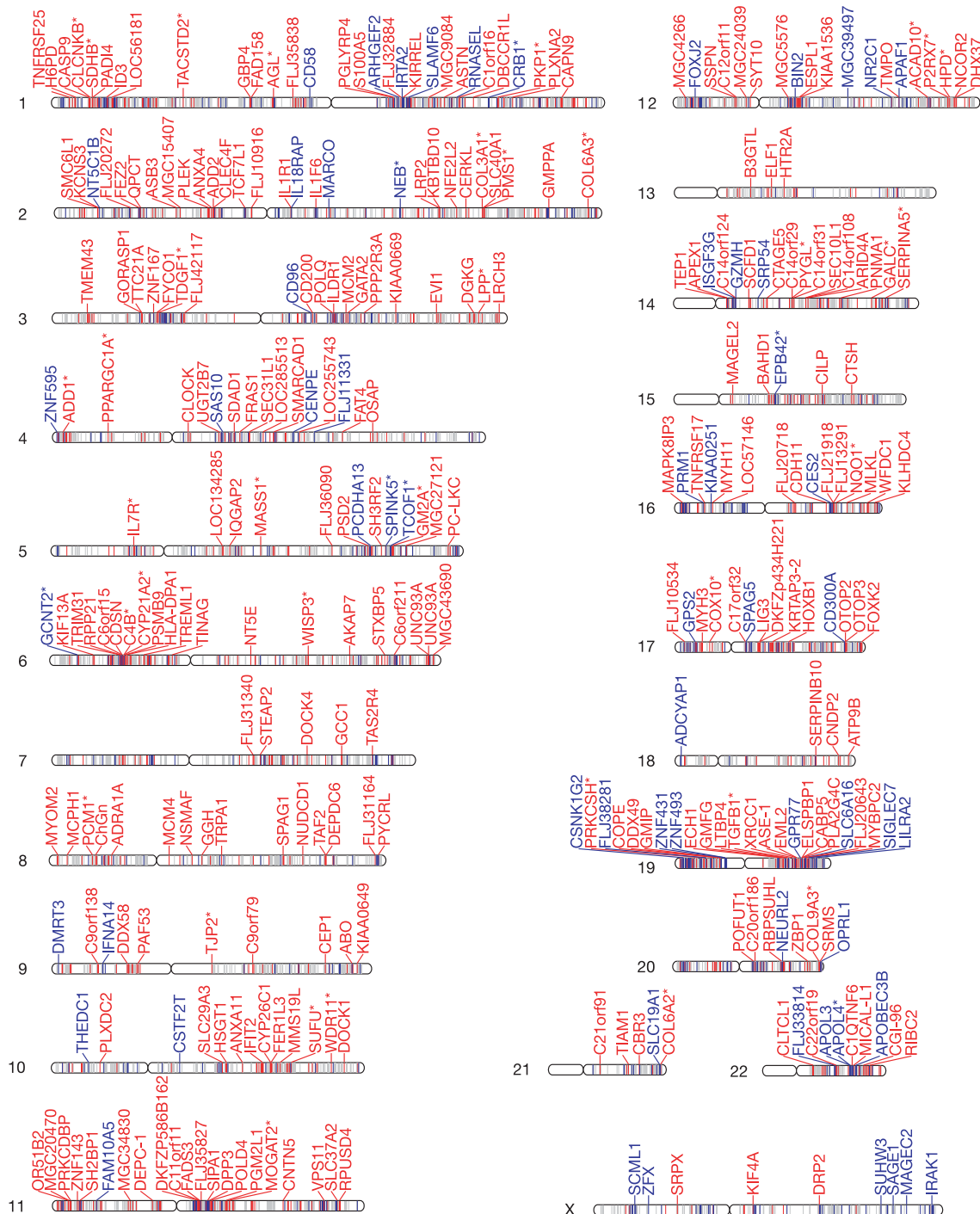


Figure 2 | A selection map of the human genome. Red bars indicate loci under negative selection and blue bars are loci under positive selection at 95% credibility level. Loci with very strong evidence of selection (>99%

credibility) are denoted by their HUGO name, and within this category genes with a morbidity entry in the OMIM database are denoted by an asterisk.

negative selection. Although we find that most of the genes in the informative data sets ($n = 4,916$) show no evidence of selection according to our methods, we do classify 813 loci as significantly negatively selected and 304 as positively selected at a 5% cutoff. Of the 50 loci identified by ref. 18 as rapidly evolving, 45 were informative about positive selection in our data set. Of these, 14 (31.1%) had more than 95% of their posterior mass above $\gamma = 0$, and 37 (82.2%) had a majority of their posterior mass above neutrality, indicating good agreement between population genetic and phylogenetic approaches for identifying rapidly evolving genes. There is a high degree of overlap in the types of genes classified as positively selected by both studies (see Table 1 and Supplementary Table 1), including defence/immunity proteins ($P = 0.00965$), gametogenesis ($P = 0.03411$), apoptosis ($P = 0.00336$) and sensory perception ($P = 0.04577$). One interesting result of our analysis is that transcription factors as a group seem to be rapidly evolving ($P < 0.0001$), with 39 out of 240 in the informative data sets (16.25%) having P^+ greater than 97.5% as compared to 9.6% of all loci in the IPS data set. Similarly, we find evidence that the categories of nuclear hormone receptors ($P = 0.00143$) and genes involved in nucleoside, nucleotide and nucleic acid metabolism ($P = 0.00467$) have an excess of rapidly evolving genes.

We also used our approach to identify loci and classes of genes that show a paucity of amino acid divergence between humans and chimpanzees, yet have moderate to high levels of amino acid polymorphism, which we believe to harbour an excess of mildly deleterious variation. For example, loci involved in actin binding ($P = 0.00013$; Supplementary Table 1) and cytoskeletal formation ($P < 0.00001$) contain an over-representation of negatively selected loci, with 36 out of 205 cytoskeletal proteins in the informative set (17.6%) having P^- greater than 97.5% (Table 1). For example, 6 out of the 9 myosin heavy chain loci exhibit a large excess of amino acid polymorphism, including non-muscle myosin (*MYH9*, $P^- > 0.983$), embryonic (*MYH3*, $P^- > 0.999$), perinatal (*MYH8*, $P^- > 0.962$) and adult skeletal (*MYH4*, $P^- > 0.946$; *MYH13*, $P^- > 0.957$) myosin as well as the smooth muscle (*MYH11*, $P^- > 0.999$) form. Other cytoskeletal proteins with excess amino acid polymorphism within human populations include myomesin 2 and 3 (*MYOM2*, *MYOM3*), dystrophin related protein 2 (*DRP2*), α - and β -adducin (*ADD1*, *ADD2*), sarcospan (*SSPN*) and scinderin (*SCIN*). These results are consistent with the fact that as a group, genes involved in cell structure and motility (Table 1) show a signature of negative or purifying selection ($P = 0.00008$), with 27 out of 176 (15.3%) loci exhibiting excess amino acid polymorphism relative to divergence.

Mutations in cytoskeletal protein-coding genes are known to cause a number of mendelian diseases and have been implicated in various complex disorders. For example, with the dystrophin gene many different types of mutation are known to cause both Duchenne and Becker types of muscular dystrophy (*DMD*, $P^- > 0.969$). Also in this set is myosin VIIA (*MYO7A*, $P^- > 0.99$), a gene implicated in Usher syndrome (1B), the most common cause of congenital deafness and blindness in developed countries. Similarly, the α - and β -adducin genes ($P^- > 0.99$ for both) are associated with hypertension and cardiovascular disease, and one known causative variant (*ADD1* G460W (refs 19, 20)) is found at moderate frequencies in both African Americans (9.7%) and European Americans (28.9%) in our sample.

Another interesting group of genes that show excess amino acid polymorphism ($P = 0.02805$) are those involved in ectoderm development, with 12 loci out of 98 exhibiting significantly elevated levels of amino acid polymorphism relative to amino acid divergence (Table 1). These include three loci in which mutations are known to cause disease: *GLI3* (polydactyly), *NOTCH3* (*Drosophila* homologue implicated in cerebral arteriopathy) and *DCC* (colorectal carcinoma). Interestingly, all three of these genes are also involved in neurogenesis according to the Panther molecular function classification. Genes involved in general vesicle transport also show

excess amino acid polymorphism as a class ($P = 0.00016$). Sixteen loci have posterior distributions with more than 95% and four with more than 99% mass above $\gamma = 0$ (for example, *COPE*, *HD*, *KIF4A*, *SEC31L1* and *STX11*). The most familiar of these is *HD*—the gene that causes Huntington's disease.

To quantify the strength of association between non-neutral evolution and genetic disease more formally, we undertook two analyses. The first is a logistic regression of Online Mendelian Inheritance in Man (OMIM) morbidity status (1 = disease; 0 = non-disease) on the square-root of observed species non-synonymous substitution rate $\sqrt{d_N}$ for all loci in the IPS data contained in the OMIM database of disease and non-disease genes. This approach (which is independent of the Poisson random field (PRF) analysis and summarized in Fig. 3a) yields a highly significant negative association ($\beta \sqrt{d_N} = -7.6941$; $\beta_0 = -0.4555$) between rate of amino acid evolution and disease status for moderate to highly polymorphic loci (likelihood-ratio test (LRT) = 14.8; $P_{\chi^2_1} = 1.1 \times 10^{-4}$; $n = 1,438$ loci), suggesting that mendelian disorders may have discernable darwinian consequences. A second test of the same hypothesis, a Mann–Whitney *U*-test comparing disease versus non-disease loci, also indicates a marginally significant excess of genes with high P^- among genes contributing to mendelian disease ($P = 0.0374$; see Fig. 3b).

The results of analysis of human polymorphism at this scale may help to guide our thinking about human evolution at a broad level, and they also help sharpen our targets for exploration of the genetic basis for medical conditions. Examination of Figs 1 and 2 shows that only a small minority of non-neutral genes are facing positive selection. In this group are genes that might be responsible for

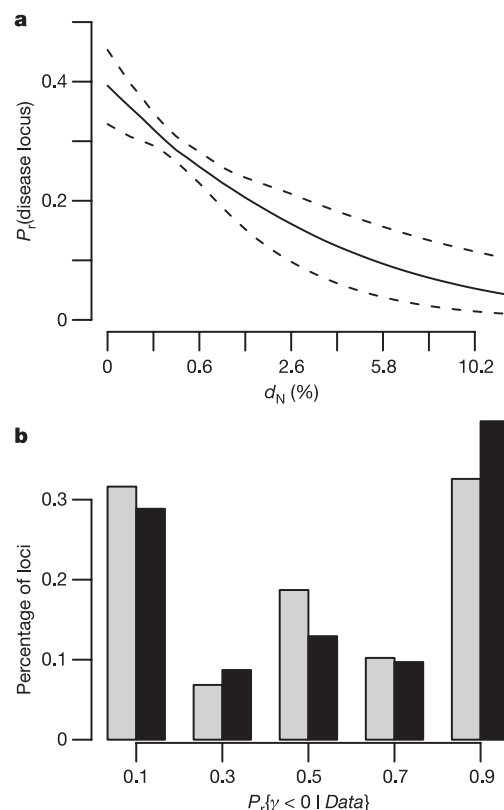


Figure 3 | Association between negative selection and disease. **a**, Results of a logistic regression analysis of OMIM status (disease versus non-disease) in the IPS data set as a function of the per site amino acid substitution rate. 95% confidence intervals are found via non-parametric bootstrap re-sampling. **b**, Histograms for $P^- = P\{\gamma < 0 | \text{Data}\}$, the probability that non-lethal mutations at a gene are negatively selected, across disease (black bars) and control (non-disease; grey bars) genes.

adaptive differences between humans and chimpanzees, or for which molecular evolution might be driven by pathogen pressure. Each of these genes is a candidate for a deeper analysis of the evolutionary past of our species. Most genes in Fig. 2 show a signature of weak negative selection, a feature that is detected only if there are segregating variants that have deleterious consequences. This implies that this group constitutes a set of genes of keen interest in medical genetics. Notably, a substantial portion of these genes is also already known to be mutable to a mendelian disorder. These genes remain prime candidates in our quest for understanding the genetic basis for a diverse array of complex disorders that show weak familial tendencies.

METHODS

Sequencing and bioinformatics. Celera Genomics applied exon-specific PCR amplification to 20,362 loci in 39 humans and one male chimpanzee in order to obtain sequence variants in these regions (details are provided in the Supplementary Methods). Of these, 14,032 genes shared at least 95% sequence identity to a unique accession in the NM series of the NCBI Reference Sequence (RefSeq) 9.0 database (3/1/2005) and mapped onto build 34 of the human genome at the same genomic location as the accession (mapping was done using BLAT²¹ v. 29). The loci sequenced by Celera were aligned with the best hit in RefSeq. Regions of the alignment that were not in the RefSeq portion of the alignment were omitted from further analyses as were loci where the chimpanzee sequence had an internal stop codon. In order to exclude paralogous genes, we restricted our selection analysis to loci with $d_s < 0.1$. Of the remaining 11,624 genes, 8,292 had at least one non-synonymous SNP or fixed difference, and could thus be used for further study using the mkprf method (McDonald–Kreitman analysis using Poisson random field; <http://cbsuapps.tc.cornell.edu/mkprf.aspx>). We also define two sets of informative loci ($n = 3,277$) with at least four variable non-synonymous sites in the alignment (that is, $P_N + D_N \geq 4$; IPS) as well as a set with at least two variable non-synonymous sites that are informative only about negative selection ($n = 6,033$; INS). Loci with less than four amino acid variable sites in the alignment cannot be informative about positive selection, and loci with 0 or 1 amino acid variable sites cannot inform a hypothesis regarding weak negative selection. The reason for this is that we make a conservative assumption that some (unknown) fraction of mutations are lethal. While inferences of selection based only on variable sites in the alignment reduces power, it does not require *a priori* assumptions about the distribution of selective effects among moderately to highly deleterious mutations.

Statistics. To maximize our power to identify non-neutrally evolving loci, we pool genomic polymorphism and divergence data using the approach of ref. 9 who used a Poisson random field setting⁴ for bayesian population genetic inference of selection. The chief advantage of the mkprf method is an increase in power to detect selection without an increase in type I error as long as per locus mutation rates are low⁹. One slight modification of the model used here is that we set a gaussian prior distribution for γ_i with mean 0 and an arbitrary standard deviation of 8 (as opposed to a hierarchical model as in ref. 9). This simplification is made so that the marginal posterior distributions of the selection coefficients are conditionally independent of one another and can be pooled for further analysis. Full details of the likelihood function used here can be found in Supplementary Data 1. Briefly, our method estimates posterior distributions of genomic parameters such as the species divergence time between humans and chimpanzees ($\tau = \text{number of generations}/2N_e$, where N_e is the effective population size of humans) using all synonymous SNPs and synonymous fixed differences in the sample. Conditional on genomic parameters, posterior distributions for individual gene parameters are informed only by the non-synonymous cell entries in a conventional McDonald–Kreitman table (that is, P_N and D_N). Evidence in our model for non-neutral evolution at a particular gene comes from the equal tail credibility intervals (a form of bayesian confidence intervals) on the selection parameter for the locus. That is, if the whole of the CI for γ_i is above 0, there is strong evidence that the locus i has an excess of amino acid fixed differences given the observed level of amino acid variability within humans at the gene, and is therefore likely to be rapidly evolving due to positive darwinian selection. Similarly, if the whole of the CI for γ_i is below 0, there are too few fixed amino acid differences in gene i ; this we interpret as evidence that the locus is subject to negative selection, and that many amino acid polymorphisms at the locus are weakly deleterious. An alternative explanation is that the locus is subject to strong balancing selection (or over-dominance) that prevents fixation of non-lethal amino acid mutations between species. Recent analytical and simulation work^{22,23} suggests that over-dominance

needs to be very strong for the McDonald–Kreitman cell entries to deviate in the direction of excess amino acid variation relative to amino acid fixed differences (such as in self-incompatibility alleles²⁴). In fact, under weak over-dominance, the McDonald–Kreitman cell entries show the same deviation (that is, an excess of amino acid fixed differences relative to polymorphism) as under positive selection²³.

Received 24 April; accepted 14 September 2005.

- Kimura, M. *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, 1983).
- Hudson, R. R., Kreitman, M. & Aguadé, M. A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–159 (1987).
- McDonald, J. H. & Kreitman, M. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**, 652–654 (1991).
- Sawyer, S. A. & Hartl, D. L. Population genetics of polymorphism and divergence. *Genetics* **132**, 1161–1176 (1992).
- Eyre-Walker, A. & Keightley, P. D. High genomic deleterious mutation rates in hominids. *Nature* **397**, 344–347 (1999).
- Fay, J. C., Wyckoff, G. J. & Wu, C. I. Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. *Nature* **415**, 1024–1026 (2002).
- Sawyer, S. A., Kulathinal, R. J., Bustamante, C. D. & Hartl, D. L. Bayesian analysis suggests that most amino acid replacements in *Drosophila* are driven by positive selection. *J. Mol. Evol.* **57** (suppl. 1), S154–S164 (2003).
- Smith, N. G. & Eyre-Walker, A. Adaptive protein evolution in *Drosophila*. *Nature* **415**, 1022–1024 (2002).
- Bustamante, C. D. *et al.* The cost of inbreeding in *Arabidopsis*. *Nature* **416**, 531–534 (2002).
- Nordborg, M. *et al.* The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol.* **3**, e196 (2005).
- Halushka, M. K. *et al.* Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nature Genet.* **22**, 239–247 (1999).
- Cargill, M. *et al.* Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nature Genet.* **22**, 231–238 (1999).
- Stephens, J. C. *et al.* Haplotype variation and linkage disequilibrium in 313 human genes. *Science* **293**, 489–493 (2001).
- Livingston, R. J. *et al.* Pattern of sequence variation across 213 environmental response genes. *Genome Res.* **14**, 1821–1831 (2004).
- Sunyaev, S. *et al.* Prediction of deleterious human alleles. *Hum. Mol. Genet.* **10**, 591–597 (2001).
- Williamson, S. *et al.* Simultaneous inference of selection and population growth from patterns of variation in the human genome. *Proc. Natl Acad. Sci. USA* **102**, 7882–7887 (2005).
- Barrier, M., Bustamante, C. D., Yu, J. & Purugganan, M. D. Selection on rapidly evolving proteins in the *Arabidopsis* genome. *Genetics* **163**, 723–733 (2003).
- Nielsen, R. *et al.* A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* **6**, e170 (2005).
- Manunta, P. *et al.* Alpha-adducin polymorphisms and renal sodium handling in essential hypertensive patients. *Kidney Int.* **53**, 1471–1478 (1998).
- Morrison, A. C., Bray, M. S., Folsom, A. R. & Boerwinkle, E. ADD1 460W allele associated with cardiovascular disease in hypertensive individuals. *Hypertension* **39**, 1053–1057 (2002).
- Kent, W. J. BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
- Weinreich, D. M. & Rand, D. M. Contrasting patterns of nonneutral evolution in proteins encoded in nuclear and mitochondrial genomes. *Genetics* **156**, 385–399 (2000).
- Williamson, S., Fedel-Alon, A. & Bustamante, C. D. Population genetics of polymorphism and divergence for diploid selection models with arbitrary dominance. *Genetics* **168**, 463–475 (2004).
- Ioerger, T. R., Clark, A. G. & Kao, T. H. Polymorphism at the self-incompatibility locus in Solanaceae predates speciation. *Proc. Natl Acad. Sci. USA* **87**, 9732–9735 (1990).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. Thornton and B. Payseur for suggestions during the analysis. Some of the analysis was supported by NIH grants to C.D.B., R.N. and A.G.C. We also acknowledge the help of J. Pillardy and the Cornell University Theory Center Computational Biology Service Unit.

Author Contributions S.G., D.M.T., D.C., T.J.W., J.J.S., M.D.A. and M.C. conceived, designed and performed the experiments. C.D.B., A.F.-A., A.G.C., S.W., R.N. and M.J.H. analysed the data.

Author Information Accession numbers for the SNP markers analysed in this study are dbSNP numbers ss48401226–ss48429818 and ss48429821–ss48431291, submitted under the handle APPLERA_GL. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.D.B. (cdb28@cornell.edu).

LETTERS

Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories

Terra D. Barnes^{1*}, Yasuo Kubota^{1*}, Dan Hu¹, Dezhe Z. Jin^{1,2} & Ann M. Graybiel¹

Learning to perform a behavioural procedure as a well-ingrained habit requires extensive repetition of the behavioural sequence, and learning not to perform such behaviours is notoriously difficult. Yet regaining a habit can occur quickly, with even one or a few exposures to cues previously triggering the behaviour^{1–3}. To identify neural mechanisms that might underlie such learning dynamics, we made long-term recordings from multiple neurons in the sensorimotor striatum, a basal ganglia structure implicated in habit formation^{4–8}, in rats successively trained on a reward-based procedural task, given extinction training and then given reacquisition training. The spike activity of striatal output neurons, nodal points in cortico-basal ganglia circuits, changed markedly across multiple dimensions during each of these phases of learning. First, new patterns of task-related ensemble firing successively formed, reversed and then re-emerged. Second, task-irrelevant firing was suppressed, then rebounded, and then was suppressed again. These changing spike activity patterns were highly correlated with changes in behavioural performance. We propose that these changes in task representation in cortico-basal ganglia circuits represent neural equivalents of the explore–exploit behaviour characteristic of habit learning.

The ability to establish habits, procedures and stereotyped behaviours brings great biological advantages to active organisms, and much evidence indicates that cortico-basal ganglia loops are critical for such learning^{4–10}. If this view were correct, changes in the activity of basal ganglia neurons should accompany changes in behaviour not only as habits and procedures are initially acquired, but also as they are changed in response to altered behavioural contexts. To test for such restructuring of basal ganglia activity, we recorded chronically with multiple tetrodes for up to 63 sessions from the sensorimotor striatum of rats undergoing consecutive acquisition, over-training, extinction and reacquisition training on a conditional T-maze task (Fig. 1, Supplementary Fig. 1 and Supplementary Table 1). The rats navigated the T-maze and turned right or left in response to auditory cues indicating whether a chocolate reward was at the left or right choice-arm of the maze (Fig. 1c). This task requires trial-and-error learning, in which initial ‘exploration’ of the environment over successive trials leads, with successful learning, to ‘exploitation’, in which correct choices are made consistently¹¹. Performance accuracy increased during acquisition and was at or near asymptote during over-training (Fig. 1d). Accuracy then steadily deteriorated during extinction training, when reward was reduced ($n = 4$) or withheld entirely ($n = 3$), but recovered rapidly during retraining after extinction. Running times similarly fell, rose and fell (Fig. 1e, f).

As these behavioural changes occurred, the spiking of striatal neurons became redistributed across task time (Fig. 2, Supplementary Figs 2 and 3). We focused on the spike activity of neurons classified as striatal projection neurons, which directly participate in

cortico-basal ganglia loop processing¹² (Fig. 1a, Supplementary Fig. 1 and Supplementary Methods). At the start of acquisition training, the spike responses of the task-responsive projection neurons, as a group, occurred throughout the maze runs (Fig. 2a). By the time that the learning criterion had been met, however, the strongest per-unit firing occurred near the start and end of the runs. This progressive concentration of spike activity continued during the over-training period, even though behavioural performance had reached near-asymptotic values. In addition, early activity advanced from the time of locomotion onset towards the waiting period after the warning cue, and late activity shifted from around goal-reaching to around the end of turning (Fig. 2a, c, and Supplementary Figs 2 and 3).

These acquired spiking patterns were largely reversed during the extinction period (Fig. 2a). Mid-trial firing increased, and the temporal shifts, particularly for the early activity, reversed. When

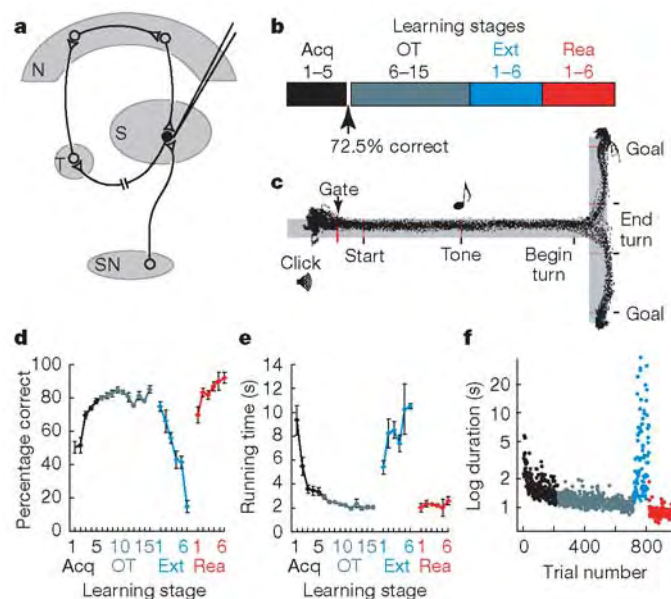


Figure 1 | T-maze task and behavioural learning. **a**, Simplified cortico-basal ganglia circuit, indicating recording of striatal projection neurons. N, neocortex; S, striatum; T, thalamus; SN, substantia nigra. **b**, Training stages (acquisition (Acq, black), 1–5; over-training (OT, grey), 6–15; extinction (Ext, blue), 1–6; reacquisition (Rea, red), 1–6; described in Supplementary Methods). **c**, Run trajectories for one over-training session. **d**, **e**, Average percentages of correct responses (**d**) and average per-trial running times (**e**). Error bars represent s.e.m.. **f**, Trial-by-trial running times for the interval between tone onset and turn onset by a rat during successive training. Each dot represents performance in one trial.

¹Department of Brain and Cognitive Sciences and the McGovern Institute for Brain Research, Massachusetts Institute of Technology, 43 Vassar Street, 46-6133, Cambridge, Massachusetts 02139, USA. ²Department of Physics, Pennsylvania State University, 0104 Davey Laboratory, University Park, Pennsylvania 16802, USA.

*These authors contributed equally to this work.

the reacquisition period then was initiated by returning the reward at the end of each correct run, there was another sudden shift in the spike patterns, producing reduced mid-trial firing and a temporal advance of start activity resembling that seen during initial acquisition (Fig. 2a, and Supplementary Figs 2–5).

To estimate the randomness of the population spike activity across the entire trial time, we calculated the entropy of the average per-neuron firing across learning stages (Supplementary Methods). The

entropy fell sharply during acquisition, rebounded during extinction, and fell again during reacquisition (Fig. 2e), in the absence of significant changes in average per-trial firing rates (Supplementary Fig. 6). The changes in the spike patterns were highly correlated with the changes in behavioural accuracy (Fig. 2f, g).

Remarkably, we found equally striking lability in the spiking patterns of the striatal projection neurons that lacked detectable phasic peri-event activity during the task (Fig. 2b). Some of these

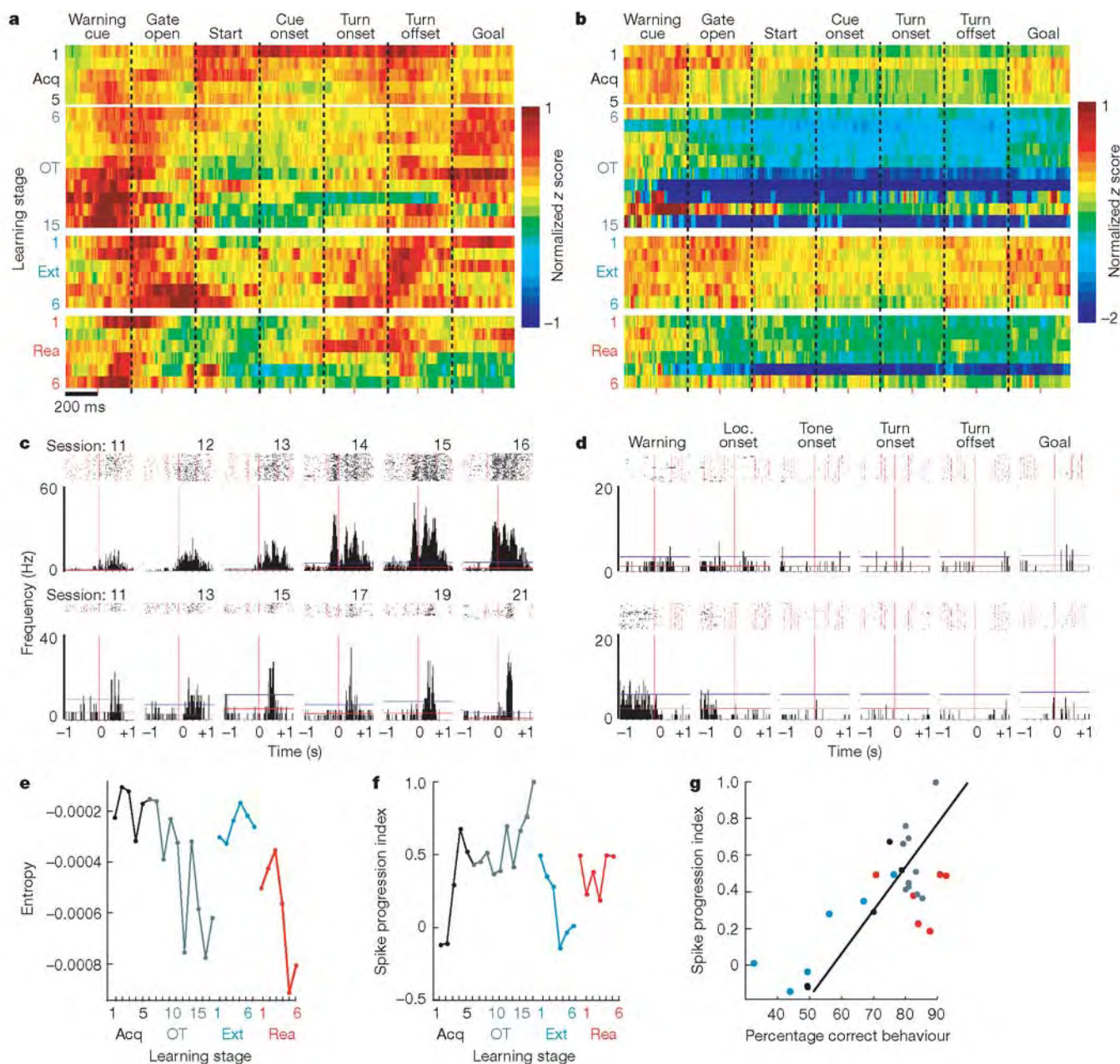


Figure 2 | Plasticity in spike activity patterns of striatal projection neurons. **a, b,** Average activity of units classified as task-responsive (**a**) and non-task-responsive (**b**) neurons plotted in 10-ms bins as z scores normalized for each neuron relative to that neuron's baseline activity, according to pseudo-colour scales shown at the right, with one row per training session. Plots show ± 200 -ms time windows around task events, abutted in the order of occurrence within a trial. **c,** Peri-event time histograms (PETHs, ± 1 -s window) for units recorded on consecutive days at single sites (putative single units), illustrating strengthening and time-shift in responses around locomotion onset over 6 consecutive sessions (top) and sharpening of phasic responses at turn onset over 13 sessions (bottom). Horizontal lines indicate mean pre-trial baseline firing rates (red) and two

standard deviations above the mean (blue, threshold for task-related activity). **d,** Typical PETHs for neurons lacking in-trial phasic peri-event activity ('non-task-responsive' units). **e,** Entropy of per-trial spike activity of task-responsive units calculated for each training stage. **f,** Spike progression index (SPI) illustrating correlation between per-trial spike activity of task-responsive projection neurons at each training stage and the neural activity at the last stage of over-training. **g,** Significant correlation ($r = 0.74$, $P < 0.0001$) between SPI and progressive changes in percentage correct behaviour during training. Black and grey, acquisition and over-training, $r = 0.82$, $P = 0.0002$; blue, extinction, $r = 0.87$, $P = 0.02$; red, reacquisition, $r = 0.09$, $P = 0.87$.

non-task-responsive neurons fired at low rates both in task and out of task, and some fired more out of task than in task (Fig. 2d). The in-task activity of these neurons dwindled during acquisition and then nearly ceased. Yet, on the first day of extinction, the average per-neuron firing of these neurons returned to pre-training levels and remained elevated. Then, when reacquisition training began, their activity declined sharply. These abrupt shifts were evident whether the activity of the neurons during the task was classified with respect to pre-trial baseline firing (Fig. 2b) or relative to in-trial activity (Supplementary Fig. 2)

To determine whether the tuning of task-related responses changed during learning, we measured multiple parameters (for example, height, width, peri-event peak timing) of the phasic spike responses to specific task events detected by a slope threshold (Supplementary Methods). None of these was altered during learning. By contrast, we found large-scale changes in the proportion of spikes per entire trial run that occurred within phasic responses (Fig. 3a, b). This proportion tripled during acquisition, fell abruptly during extinction and abruptly rose again during reacquisition (Fig. 3b). The number of phasic responses also changed successively (Fig. 3c). Reinforcing these redistributions of spike activity, the proportions of task-responsive projection neurons responding to different task events also progressively emphasized⁷, de-emphasized and re-emphasized the beginning and end of the task (Fig. 3d, and Supplementary Fig. 6). Notably, the sharpening of phasic responses during acquisition held not only for those occurring in the early and late parts of the task runs in which overall spiking increased, but also for responses occurring in the middle parts of the runs in which spike activity decreased (Supplementary Fig. 6). This result indicates that even when fewer neurons responded, some 'expert' responders with sharpened responses developed in the striatum as the task was acquired. This property, too, was subject to reversal and reappearance during subsequent extinction and reacquisition training.

Both the increase in concentration of spikes within phasic peaks during acquisition and the redistribution of spikes across run times had the effect of reducing the spread of spiking across trial performance time as the rats learned the task. We looked for, but did not

observe, significant changes in the variability of firing rates within peri-event or phasic-response windows across learning. However, we found major changes in the entropy (Fig. 2e) and in the variance (Supplementary Fig. 6) of spiking activity across the entire maze runs. Changes in spike distribution within the time frame of the entire procedural performance thus represented the key modulation of spike variability that we detected during learning.

Taken together, our findings show that per-trial spike distributions, response tuning and task selectivity were dynamically reconfigured as the procedural behaviour was acquired, extinguished and reacquired. Composite neural activity scores based on these factors were highly correlated with both behavioural accuracy and running times, especially during acquisition and extinction (Fig. 4, Supplementary Fig. 7 and Supplementary Methods). Restructuring of the day-by-day neural activity patterns in the 'fast learners' ($n = 5$) but not in the 'slow learners' ($n = 2$) early during acquisition (Supplementary Fig. 8) favoured a primary correspondence between the evolution of the neural restructuring over time and associative learning. The acquired patterns were detectable in both correct and incorrect trials (Supplementary Fig. 9), however, so that the ensemble patterns were not tied to performance in individual trials.

It has been proposed that the basal ganglia promote variability in behaviour during trial-and-error learning (exploration) and serve to evaluate behavioural changes to promote the acquisition of optimal behaviour (exploitation)^{9,10,13,14}. Our findings suggest that there might be a direct neural analogue to this explore-exploit behaviour in the firing patterns of projection neurons in the sensorimotor striatum. We demonstrate two fundamental changes in the spike activity of striatal projection neurons during procedural learning.

First, there was a global modulation of the firing of projection neurons. Early in training, the spike activity of the task-responsive population was spread throughout task time, as though all task events were salient (neural exploration). Even neurons without detectable phasic task-responsive activity fired at low rates during the task. Then, with continued training, this widespread spiking of the task-responsive population diminished, and their spike activity became focused (neural exploitation). At the same time, the non-task-responsive population fell silent, further reducing the task-irrelevant firing of the total projection neuron population. These changes in firing thus altered the distribution of striatal output neuron firing across the actual time-frame of the behaviour to be learned (the entire task run time). The reversal of the acquired task-related patterning during extinction and its reappearance in reacquisition fits the idea of increased neural exploration in the new extinction context and then a return to neural exploitation in the reinstated original context during reacquisition^{15–17}. The vivid

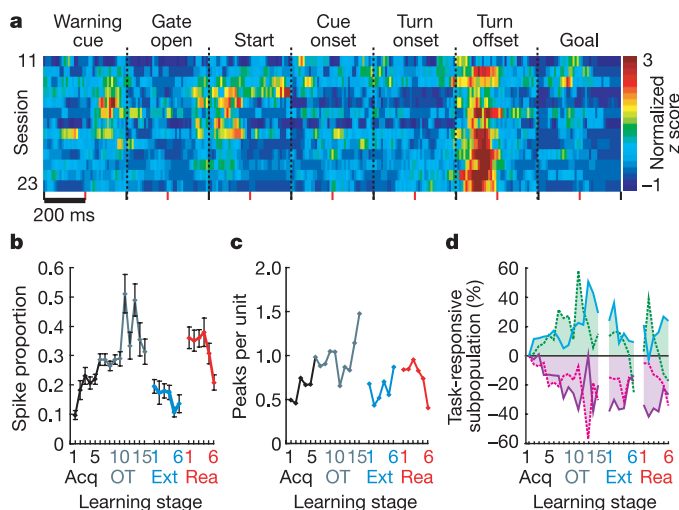


Figure 3 | Multiple changes in projection neuron activity in the sensorimotor striatum during acquisition, extinction and reacquisition training. **a**, Unit activity at a single site recorded over 13 sessions. Each row represents a session. **b**, Proportions of spikes concentrated in phasic responses in each trial, averaged for each session. Error bars indicate s.e.m. **c**, Average numbers of phasic responses per unit. **d**, Percentages of task-responsive projection neurons with responses at warning cue (solid blue line), goal reaching (dotted green line), locomotion onset (solid purple line) or turning (dotted magenta line). Values are plotted relative to the first training stage.

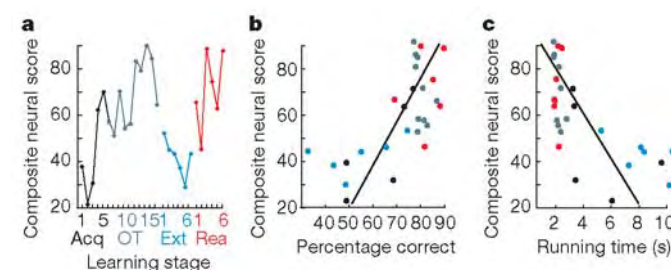


Figure 4 | Striatal neural activity predictive of behavioural performance. **a**, Composite neural scores based on weighted neural measures at trial start (normalized per-neuron firing rates during the ± 200 -ms interval around the warning cue, proportions of warning cue-responsive neurons, and proportions of spikes within phasic warning cue responses). **b**, Significant correlation between the composite scores (shown in **a**) and actual behavioural accuracy for each training stage (colour-coded as in **a**; $r = 0.69$, $P < 0.001$). **c**, Plot as in **b**, showing significant correlation between the composite neural scores and actual running times for each training stage ($r = -0.69$, $P < 0.001$).

modulation of the spiking of striatal projection neurons without detectable task-related activity also accords with this interpretation.

Second, in the exploitation phases of learning, ensemble firing at the start and end of the learned procedure strengthened, and sharply tuned responses of 'expert neurons' appeared. These changes indicate that early in training many candidate neurons fired, but that, with training and presumably competitive selection^{18–20}, neurons with sharply tuned responses appeared and, as a population, were tuned preferentially to respond near the start and end of the entire procedure performed. Our experiments leave open the question of where within the sensorimotor cortico-basal ganglia loop such changes were initiated. Because we recorded from striatal projection neurons, however, our findings show that such learning-related changes in neuronal firing occur as part of cortico-basal ganglia loop processing. The learning-related reduction in firing during the middle of the task time could indicate that striatal activity during this time was no longer needed for task execution, but could reflect the marking of behavioural boundaries in the process of chunking of the entire task performance⁵. These changing patterns could, in turn, reflect ongoing reorganization of cortico-basal ganglia activity^{20–23}. If so, the patterns could reflect neural representations related to the ready release of the learned behaviours by the appropriate context⁵.

Cortico-basal ganglia circuits probably act in determining, through reinforcement-based evaluation, which actions to enhance or diminish as learning proceeds^{4–6,9,10,19,20,24–30}. Viewed in the context of such selection functions, our findings indicate that dynamic neural representations in the striatum could adjust the encoded salience of task events and behavioural responses as habits are formed, lost and regained.

METHODS

The spike activity of neurons in the sensorimotor striatum was recorded chronically during behavioural training on a conditional reward-based T-maze task for 24–63 daily sessions from seven rats in which seven tetrode headstage assemblies had been implanted. Recordings began on the first day that the rats received training (about 40 trials per day) on the task, and were continued through successive acquisition training (stages 1–5), over-training (stages 6–15), extinction (stages 1–6) and reacquisition training (stages 1–6; Fig. 1b, Supplementary Table 2 and Supplementary Methods). In this task, rats learned to run down the maze and to turn right or left as instructed by auditory cues in order to receive reward. Behavioural data were acquired by means of photo-beams and a CCD camera. Neural data (32 kHz sampling) were collected by means of a Cheetah Data Acquisition System (Neuralynx Inc.). Well-isolated units accepted after cluster cutting were classified as striatal projection neurons or interneurons (Supplementary Fig. 1b–d). Behavioural and neural data were aligned by time stamps and were analysed by in-house software. The properties of both task-responsive and non-task-responsive projection neurons were analysed. Task-related responses of putative projection neurons were identified with respect to activity during a pre-trial 500-ms baseline period (threshold: 2 s.d. above baseline mean) and used to define task-responsive and non-task-responsive populations (Supplementary Methods). Unit data were analysed per neuron and per neuronal population across task events (Fig. 1c). To analyse population activity, normalized firing rates were averaged for each learning stage, and indices of spike firing patterns across learning stages were computed. The proportions of neurons with different task-related response types, the proportions of spikes that occurred within peri-event phasic responses per session, and trial-to-trial spike variability were also calculated, along with composite neural scores and measures of the entropy of neural firing. Changes in these measures were compared to changes in per cent correct performance and running times of the rats across stages of training.

Received 2 June; accepted 21 July 2005.

- James, W. *The Principles of Psychology* 104–127 (Dover, New York, 1890).
- Dickinson, A. Actions and habits: the development of behavioural autonomy. *Phil. Trans. R. Soc. Lond. B* **308**, 67–78 (1985).

- Pavlov, I. P. *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex* (ed. and transl. Anrep, G. V.) (Oxford Univ. Press, London, 1927).
- Packard, M. G. & Knowlton, B. J. Learning and memory functions of the basal ganglia. *Annu. Rev. Neurosci.* **25**, 563–593 (2002).
- Graybiel, A. M. The basal ganglia and chunking of action repertoires. *Neurobiol. Learn. Mem.* **70**, 119–136 (1998).
- Poldrack, R. A. *et al.* Interactive memory systems in the human brain. *Nature* **414**, 546–550 (2001).
- Jog, M., Kubota, Y., Connolly, C. I., Hillegaart, V. & Graybiel, A. M. Building neural representations of habits. *Science* **286**, 1745–1749 (1999).
- Yin, H. H., Knowlton, B. J. & Balleine, B. W. Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *Eur. J. Neurosci.* **19**, 181–189 (2004).
- Olveczky, B. P., Andalman, A. S. & Fee, M. S. Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. *PLoS Biol.* **3**, e153 (2005).
- Kao, M. H., Doupe, A. J. & Brainard, M. S. Contributions of an avian basal ganglia-forebrain circuit to real-time modulation of song. *Nature* **433**, 638–643 (2005).
- Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction* (MIT Press, Cambridge, Massachusetts, 1998).
- Wilson, C. J. in *The Synaptic Organization of the Brain* (ed. Shepherd, G. M.) 361–413 (Oxford Univ. Press, New York, 2004).
- Doya, K. & Sejnowski, T. J. in *Advances in Neural Information Processing Systems* Vol. 7 (eds Tesauero, G., Touretzky, D. S. & Leen, T. K.) 101–108 (MIT Press, Cambridge, Massachusetts, 1995).
- Doya, K. & Sejnowski, T. J. in *The New Cognitive Neurosciences* (ed. Gazzaniga, M. S.) 469–482 (MIT Press, Cambridge, Massachusetts, 2000).
- Bouton, M. E. Context and behavioural processes in extinction. *Learn. Mem.* **11**, 485–494 (2004).
- Routtenberg, A. & Kim, H.-J. in *Cholinergic–Monoaminergic Interactions in the Brain* (ed. Butcher, L. L.) 305–331 (Academic, New York, 1978).
- Myers, K. M. & Davis, M. Behavioral and neural analysis of extinction. *Neuron* **36**, 567–584 (2002).
- Gurney, K., Prescott, T. J. & Redgrave, P. A computational model of action selection in the basal ganglia. I. A new functional anatomy. *Biol. Cybern.* **84**, 401–410 (2001).
- Graybiel, A. M., Aosaki, T., Flaherty, A. W. & Kimura, M. The basal ganglia and adaptive motor control. *Science* **265**, 1826–1831 (1994).
- Djurfeldt, M., Ekeberg, Ö. & Graybiel, A. M. Cortex-basal ganglia interaction and attractor states. *Neurocomputing* **38–40**, 573–579 (2001).
- Frank, M. J., Loughry, B. & O'Reilly, R. C. Interactions between frontal cortex and basal ganglia in working memory: a computational model. *Cogn. Affect. Behav. Neurosci.* **1**, 137–160 (2001).
- Houk, J. C. & Wise, S. P. Distributed modular architectures linking basal ganglia, cerebellum, and cerebral cortex: their role in planning and controlling action. *Cereb. Cortex* **5**, 95–110 (1995).
- Doya, K. Metalearning and neuromodulation. *Neural Netw.* **15**, 495–506 (2002).
- Montague, P. R., Hyman, S. E. & Cohen, J. D. Computational roles for dopamine in behavioural control. *Nature* **431**, 760–767 (2004).
- Tanaka, S. C. *et al.* Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nature Neurosci.* **7**, 887–893 (2004).
- Reynolds, J. N. J., Hyland, B. I. & Wickens, J. R. A cellular mechanism of reward-related learning. *Nature* **413**, 67–70 (2001).
- Mink, J. W. The basal ganglia: focused selection and inhibition of competing motor programs. *Prog. Neurobiol.* **50**, 381–425 (1996).
- McClure, S. M., Berns, G. S. & Montague, P. R. Temporal prediction errors in a passive learning task activate human striatum. *Neuron* **38**, 339–346 (2003).
- Barto, A. G. in *Models of Information Processing in the Basal Ganglia* (eds Houk, J., Davis, J. & Beiser, D.) 215–232 (MIT Press, Cambridge, Massachusetts, 1995).
- Dickinson, A. & Balleine, B. W. Motivational control of goal-directed action. *Anim. Learn. Behav.* **22**, 1–18 (1994).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank H. F. Hall, P. A. Harlan and C. Thorn for their help. This work was funded by the National Institutes of Health and the Office of Naval Research.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.M.G. (graybiel@mit.edu).

LETTERS

Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution

Elodie Ghedin¹, Naomi A. Sengamalay¹, Martin Shumway¹, Jennifer Zaborsky¹, Tamara Feldblyum¹, Vik Subbu¹, David J. Spiro¹, Jeff Sitz¹, Hean Koo¹, Pavel Bolotov², Dmitry Dernovoy², Tatiana Tatusova², Yiming Bao², Kirsten St George³, Jill Taylor³, David J. Lipman², Claire M. Fraser¹, Jeffery K. Taubenberger⁴ & Steven L. Salzberg^{1,5}

Influenza viruses are remarkably adept at surviving in the human population over a long timescale. The human influenza A virus continues to thrive even among populations with widespread access to vaccines, and continues to be a major cause of morbidity and mortality^{1,2}. The virus mutates from year to year, making the existing vaccines ineffective on a regular basis, and requiring that new strains be chosen for a new vaccine. Less-frequent major changes, known as antigenic shift, create new strains against which the human population has little protective immunity, thereby causing worldwide pandemics. The most recent pandemics include the 1918 'Spanish' flu, one of the most deadly outbreaks in recorded history, which killed 30–50 million people worldwide, the 1957 'Asian' flu, and the 1968 'Hong Kong' flu³. Motivated by the need for a better understanding of influenza evolution, we have developed flexible protocols that make it possible to apply large-scale sequencing techniques to the highly variable influenza genome. Here we report the results of sequencing 209 complete genomes of the human influenza A virus, encompassing a total of 2,821,103 nucleotides. In addition to increasing markedly the number of publicly available, complete influenza virus genomes, we have discovered several anomalies in these first 209 genomes that demonstrate the dynamic nature of influenza transmission and evolution. This new, large-scale sequencing effort promises to provide a more comprehensive picture of the evolution of influenza viruses and of their pattern of transmission through human and animal populations. All data from this project are being deposited, without delay, in public archives.

The genomes reported here comprise the initial results from the Influenza Genome Sequencing Project, a partnership between the US National Institute of Allergy and Infectious Diseases and collaborators from around the world⁴, whose goal is to sequence the genomes of thousands of influenza virus isolates. This study is the first, to our knowledge, to attempt to sequence strains that were not pre-selected for particular virulence or other unusual characteristics, and should therefore provide a relatively unbiased view of influenza virus strains in the population. Here we focus on a collection from New York State spanning several years, and subsequent studies will focus on samples from multiple, distant geographical sources across a longer time span. As our analysis shows, even within a geographically constrained set of isolates, we have found surprising genetic diversity, indicating that the reservoir of influenza A strains in the human population—and the concomitant potential for segment exchange between strains—may be greater than was previously suspected.

The genome of the influenza A virus (family Orthomyxoviridae) consists of eight single-stranded negative sense RNA molecules spanning approximately 13.5 kilobases (kb). The segments range in length from 890 to 2,341 nucleotides and encode a total of 11 proteins. Although a large number of partial influenza A virus sequences now exist in the public archives (for example, GenBank), relatively few complete genomes are available. In part this is due to the technical difficulty of constructing an efficient sequencing pipeline for an RNA-based organism. The bulk of the public data on influenza comprises short fragments from the haemagglutinin (HA) or neuraminidase (NA) segments of the genome, which encode the two main surface proteins and which, it is widely believed, are the source of most of the antigenic variation in the virus. As a result of this project, the number of complete human H3N2 influenza virus genomes in GenBank has already grown from just seven genomes to over 200.

We have completely sequenced all eight segments from 207 H3N2 isolates and two H1N2 isolates. In total, the finished sequence covers 2,821,103 bases, with an average of 13,498 bases per isolate. Table 1 shows the sequencing results for all isolates, broken down by segment. The polymerase chain reaction with reverse transcription (RT-PCR)-based sequencing strategy produces an average of 5.6 sequencing reads covering each nucleotide, as shown in Table 1. The average Phred quality value⁵ was 33, which at 5.6-fold coverage corresponds to an error rate of 3.2×10^{-19} ; however, regions of low coverage will have higher error rates. These regions can be inspected at the NCBI Assembly Archive⁶, which displays the raw data underlying every nucleotide in each genome. Note that the error rate in these genomes is likely to be considerably lower than previously sequenced influenza isolates, which in many cases reflect single-pass or two-pass sequencing. Assembly was performed using the Minimus assembler program⁷ followed by the AutoEditor program to correct erroneous bases⁸. Details of assembly and annotation are provided in the Supplementary Methods.

This is the first large-scale analysis of influenza isolates collected in a relatively unbiased manner, allowing a comprehensive look at an influenza virus population across several seasons within a constrained geographical area. Among RNA viruses, only human immunodeficiency virus (HIV) has been subjected to similar whole-genome analysis^{9,10}. In this first set of 209 genomes, we have observed multiple, novel mutational events, including point mutations, deletions and segment exchange. By carefully cataloguing these events, we can begin to get the first real picture of the rate of mutational events underlying influenza A virus evolution.

¹The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, Maryland 20850, USA. ²National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894, USA. ³Wadsworth Center, New York State Department of Health, Albany, New York 12201, USA. ⁴Department of Molecular Pathology, Armed Forces Institute of Pathology, Rockville, Maryland 20850, USA. ⁵Center for Bioinformatics and Computational Biology, University of Maryland Institute for Advanced Computer Studies, College Park, Maryland 20742, USA.

Although extensive previous research has catalogued changes in the HA and NA segments, we found new mutations in these segments as well as multiple changes in the other six segments. Some of these changes are shown in Fig. 1 and discussed below, and a comprehensive list of amino acid mutations in all eight segments is provided in the Supplementary Data.

Perhaps the most dramatic finding in our data is the discovery of an epidemiologically significant reassortment that explains the appearance, during the 2003–2004 season, of the ‘Fujian/411/2002’-like strain, for which the existing vaccine had limited effectiveness. In a recent paper¹¹, we described how phylogenetic analysis of 156 H3N2 genomes from our project revealed the clear presence of multiple, distinct clades circulating in the population. Through a reassortment event, a minor clade provided the haemagglutinin gene that later became part of the dominant strain after the 2002–2003 season. Two of our samples, A/New York/269/2003 (H3N2) and A/New York/32/2003 (H3N2), show that this minor clade continued to circulate in the 2003–2004 season, when most other isolates were reassortants. In these samples (Fig. 1, see columns for November 2003) the HA segment is clearly similar to the dominant clade, whereas the other segments all show numerous differences.

This finding illustrates not only that the influenza virus population contains multiple lineages at any given time, but also that alternate, minor lineages can contribute genetic variation to the dominant lineage, resulting in epidemiologically significant, antigenically novel strains. It is worth emphasizing that our sequence-based sampling approach—in contrast to traditional serologically based sampling—will reveal co-circulating strains even before they become antigenically novel.

Figure 1 illustrates five seasons’ worth of mutations in all proteins from the 207 H3N2 influenza virus isolates included in this study. For clarity, amino acid positions are shown only if they underwent genetic changes in at least three isolates. Each mutation is indicated by a colour shift along a row in the figure. For example, the first row shows that the amino acid in position 5 of HA1 mutated from glycine (G, shown in light green) to valine (V, shown in burgundy) in November 1999, and then back to glycine in November 2001 and afterwards, except for three isolates in January to February 2002 that show a glutamic acid (E, shown in pink) at that position. In total, 186 positions experienced at least one amino acid change.

As the figure shows, mutations appear both during and between influenza seasons. For example, HA residues 5, 33 and 92 remained unchanged from May 1999 to October 1999, and then mutated in November 1999, leading to a permanent switch for the rest of that season. Notably, multiple changes in the internal segments, including those encoding the polymerase genes (PA, PB1 and PB2), the nucleocapsid protein (NP), and two non-structural proteins (NS1 and NS2), first appeared in the 2001–2002 season and became fixed thereafter.

Data from isolates collected in the spring of 2003 provide a glimpse

of the transitional period before a major reassortment event. Many of the HA mutations that became dominant during the 2003–2004 influenza season first appeared in February 2003. Mutations to residues 155 and 156 of the HA1 domain (H155T and Q156H) show up in early 2003; these sites are accessible to antibodies and had an important role in the antigenic mismatch between the vaccine strain and the circulating viruses in the 2003–2004 season¹². A different picture emerges in the other proteins, where mutations that appear in February 2003 remain only in a few isolates. This clearly indicates that a reassortment event brought in a new HA segment during or before the spring of 2003, and subsequent data show that this reassortant strain became dominant in the 2003–2004 influenza season¹¹.

A number of important mutations found in our data may affect receptor-binding affinity and potentially increase viral replication efficiency. Studies have determined that changes in HA residues 183, 186 and 226 could affect HA receptor-binding affinity¹³, and residue positions 131, 222, 225 and 226 are important for efficient replication¹². Mutation S186G appears in circulating viruses during the 2001–2002 influenza season, along with mutation V202I, and remains in the 2003–2004 season. Mutations A131T, W222R and G225D also emerge in February 2003. The HA1 T155H and H156Q mutations in our data are accompanied by a possibly correlated mutation at residue 25 (L25I).

The neuraminidase protein has a box-shaped globular head with four catalytic sites that allow the cleavage of sialic acid linkages¹⁴. Amino acid positions important for antigenic drift have been identified for the N2 subtype¹⁴ as well as other regions likely to be involved in virus–host interactions and qualified as phylogenetically important regions¹⁵. Sequence data in our study indicate that once residue 197—an antigenic site^{14,16}—mutated from 197H to 197D early in the 1999–2000 influenza season, it was accompanied by the mutation R249K. This residue is probably not in a functional site but may be functionally compensating by maintaining the accessibility of surface residues. Residue 199 interestingly switched (E199K) for the 2003–2004 influenza season for the majority of the isolates, except for the two isolates corresponding to the minor non-reassorted clade (A/New York/269/2003 and A/New York/32/2003)¹¹.

Table 2 lists correlated mutations that may be co-mutations; that is, where there appears to be a balancing effect between two sites on the same protein, or between a site on an internal protein and one on a surface protein. The best example is seen for T392 in NA, which is present in the same eight isolates (appearing in 2001–2003) where there is an I463 mutation in PB2 (see Fig. 1).

The fact that the minor Fujian-like clade has donated its HA to the previously dominant strain rather than itself becoming the dominant circulating virus indicates that there may be important amino acid co-substitutions in the other proteins essential for viral fitness¹¹. When comparing the NA and internal proteins of the dominant circulating major clade present during the 2003–2004 influenza season with the previous dominant clade, there are a few

Table 1 | Sequencing results for 209 complete genomes of H3N2 and H1N2 human influenza A viruses

Segment	Length (nt)	Size of coding region (nt)	Total finished sequence (nt)	Finished sequence per segment (nt)	Coding completeness (%)	Average segment coverage*
PB1	2,341	2,274	483,932	2,315	100	5.5
PB2	2,341	2,280	482,950	2,311	100	5.3
PA	2,233	2,151	463,810	2,219	100	5.9
HA	1,762	1,701	364,012	1,742	100	5.8
NP	1,565	1,497	324,681	1,553	99.9	5.4
NA	1,466	1,410	303,410	1,452	99.7	5.9
M	1,027	982	213,806	1,023	100	5.2
NS	890	838	184,502	883	100	5.7
Total	13,625	13,133	2,821,103	13,498	99.95	5.6

nt, nucleotide.

* Coverage is defined as the average number of individual sequencing reads covering each nucleotide of finished sequence.

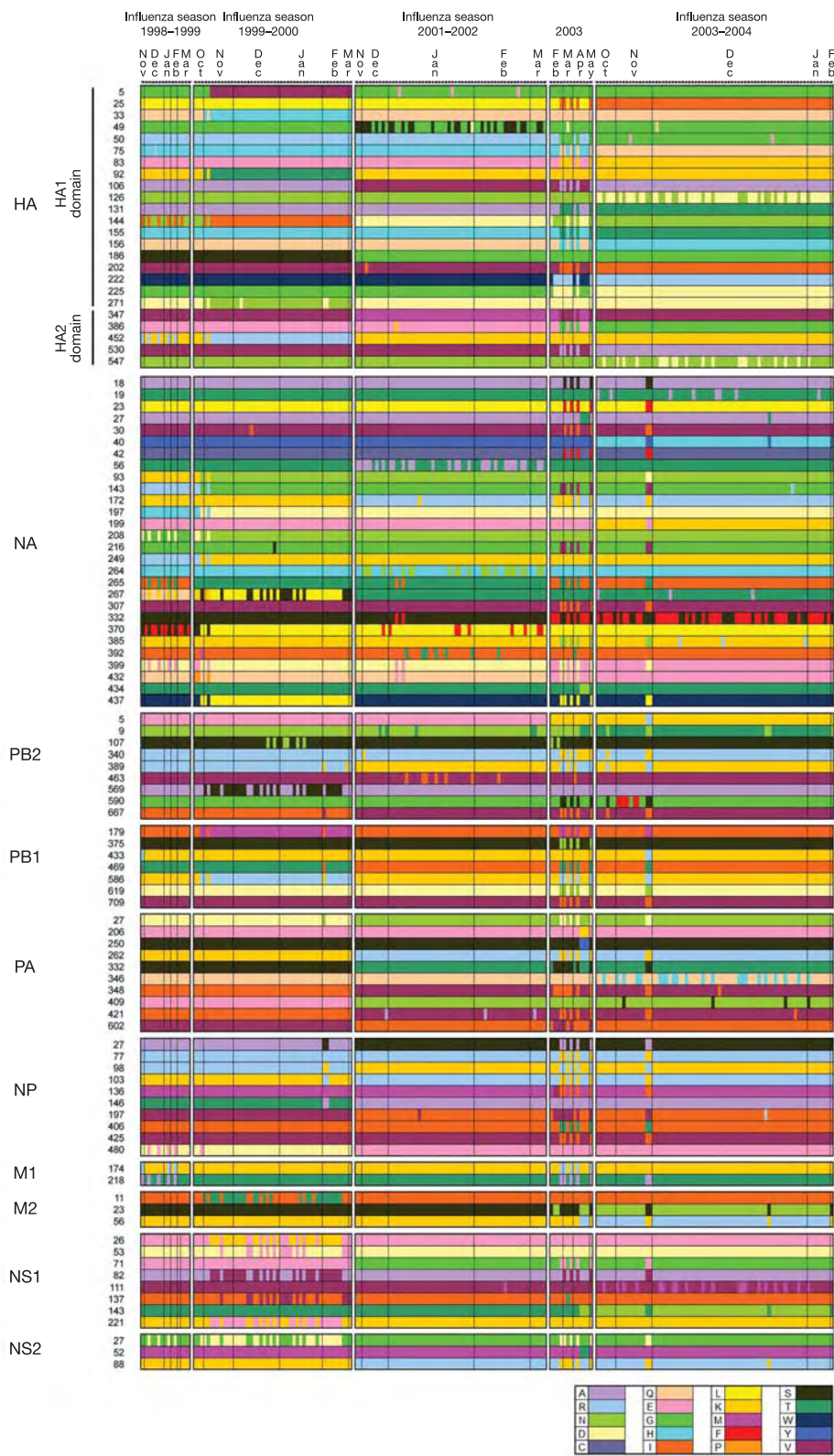


Figure 1 | Sites with genetic changes across the ten main proteins in 207 influenza A viruses. Each row represents a single amino acid position in one protein. Amino acids (single-letter abbreviations are used) are colour-coded as shown in the key, so that mutations can be seen as changes in colour when scanning from left to right along a row. For simplicity, only amino acids that showed changes in at least three isolates are shown. Each column represents a single isolate, and columns are only a few pixels wide in order to

display all 207 H3N2 isolates in this figure. Isolates are ordered along the columns chronologically according to the date of collection; boundaries between influenza seasons are indicated by gaps between columns. A more detailed version of this figure, showing positions that experienced any amino acid change and showing identifiers for the isolates in each column, is available as Supplementary Fig. 1.

Table 2 | Correlated mutations within and between influenza virus proteins

Correlated mutations within segments						
HA*	NA	PB2	PB1	PA	NP	NS1
N30-S400	L45-D401	R389-I667	I179-K586	D27-K262	A27-R98	Q25-N127
S30-T400	P45-G401	K389-V667	M179-R586	N27-R262	S27-K98	K25-D127
K269-V476	H197-R249		D619-V709	E206-S250	1406-V425	E26-K221
R269-A476	D197-K249		N619-I709	K206-Y250	T406-I425	K26-E221
D275-L304					A451-S482	
I144-R252					T451-N482	
					I136-I425	
Correlated mutations of surface and internal protein residues						
HA*-NA	HA*-M1	HA*-PA	NA-PB2			
G49-T56	F159-K208	G275-N359	T392-I463			
S49-A56		D275-D359				
N547-D147						
D547-N147						

*HA residues are counted from the first amino acid of the HA1 domain.

substitutions that appear to be unique. NA Y40H and E199K, for example, occur in regions potentially affecting virus–host interactions¹⁵. On the basis of the NA alignments performed using data from the LANL influenza database (data not shown), the E199K mutation is also seen in southeast Asian isolates that were collected during the 2003–2004 influenza season when the Fujian-like variant was the dominant circulating virus in that region. Over the rest of the NA protein, however, the southeast Asian isolates resemble the non-reassortant North American clade B¹¹. This indicates that this clade may have had deficiencies in certain residues that affected its ability to become the dominant virus. There is, unfortunately, no publicly available data for these proteins from the southeast Asian isolates with which to do a comparative analysis to the North American major and minor clades.

Recent reports^{17–19} have described a newly discovered protein, known as PB1-F2, encoded by a shifted reading frame in the PB1 gene. The data presented here more than double the total number of complete PB1 segments in the public archives, and we found that the PB1-F2 open reading frame is preserved in 206 out of 209 PB1 genes. In most cases (180 out of 206) the protein's length is 90 amino acids, but it is 87 amino acids in 23 isolates and 80 amino acids in three isolates. In three cases, an in-frame stop truncates the predicted protein after 11 amino acids. The translations of PB1-F2 for all 209 isolates are provided in the Supplementary Data.

Our project includes two clear examples of segment exchange between H1N1 and H3N2 viruses, both of which are H1N2 serotypes. In both isolates from our collection, only the haemagglutinin segment was exchanged, and these appear to be descendants of a human–swine recombinant, as has been reported previously²⁰. Although segment exchange has been reported before^{21,22}, no accurate data on the frequency of these events have been collected. Our observation of three events (two exchanges between different H3N2 clades¹¹, and one exchange between H3N2 and H1N1) in 209 samples may provide an initial baseline for future estimates.

The Influenza Genome Sequencing Project is currently being expanded to include avian influenza, in an effort to establish how often these strains cross the species barrier and move into the human population. One possible cause for influenza pandemics is the mixing through reassortment of an avian influenza strain with a human strain via co-infection of a single host²³. Recent reports of transmissions of avian influenza virus to humans^{24,25} have raised concerns that a new pandemic might emerge²⁶. Despite the importance of the threat that influenza presents, no previous effort has been made to study its complete genome on a large scale. The protocols described here are being generalized to include large numbers of avian influenza isolates that, like the genomes reported here, will be deposited immediately in public archives.

METHODS

All sequence data used in this study are available from GenBank, and also via a project page at <http://www.tigr.org/flu>. In addition, all 209 genomes and GenBank accession numbers are available as a single file in the Supplementary Data.

All samples for this study were collected by the Virus Reference and Surveillance Laboratory of the Wadsworth Center in Albany, New York, which maintains a repository of human influenza samples dating back to 1992. Virus samples were received as part of outbreak investigations, through the reference function of the laboratory, and, since 2001, as part of a sentinel physician influenza programme. Use of the diagnostic samples in this study was approved by the New York State Department of Health Institutional Review Board.

Viral RNA isolation. Isolates were amplified in tube cultures of primary rhesus monkey kidney (pRhMK) cells before extracting 140 µl of culture supernatant. Viral RNA was extracted from clarified supernatant fluid using the Qiagen BioRobot M48 workstation with the MagAttract Viral RNA M48 kit (Qiagen).

RNA ligation. RNA was circularized overnight at 4 °C with T4 RNA ligase (Epicentre). Before the ligation step, the RNA was first treated with tobacco acid pyrophosphatase (20 U TAP in a 15-µl reaction, incubated at 37 °C for 1 h). TAP treatment is usually used to remove molecules from the 5' end of RNA, mostly plus-strand RNA. Although no such molecules are expected to be present on the influenza genomic RNA segments, ligation was more efficient with this treatment than without. The circularized RNA was cleaned again with the RNeasy Mini kit (Qiagen).

RT-PCR and sequencing. The first step in the high-throughput sequencing pipeline uses reverse transcription followed by polymerase chain reaction amplification to generate overlapping DNA amplicons covering each segment of the influenza virus genome. Overlapping primers were designed approximately every 200–250 nucleotides along the genome; degenerate primers allow the pipeline to tolerate sequence variation. In order to capture the extreme ends of each segment, we used an RNA circularization step before the RT-PCR²⁷. We then used RT-PCR to amplify a chimaeric product that contained the sequence from both ends of the segment.

Complementary DNA synthesis. RT-PCRs were performed with a OneStep RT-PCR kit (Qiagen). Ninety-five reactions were performed per RNA sample. Degenerate primers were designed based upon the alignment of selected human H3N2 sequences. For most of the segments, all full-length and nearly full-length sequences from 1980 to the present were aligned and used for primer design. For others, more stringent criteria were used in order to reduce the number of sequences in the set to a more manageable number. An M13 sequence tag was added to the 5' end of each primer to be used for sequencing (F primers: TGTAACACGACGCCAGT; R primers: CAGGAACAGCTATGACC). Eight pairs of primers were designed to span the ligated ends of each segment to capture the end sequences. Four of the reactions were analysed on an agarose gel for quality control purposes. Primer sequences are included as a separate table in the Supplementary Data.

Amplicons were prepared for sequencing by incubating them at 37 °C for 60 min with 0.5 U of shrimp alkaline phosphatase (Amersham) and 1 U of exonuclease I (Amersham) to inactivate remaining dNTPs and to digest the single-stranded primers. The enzymes were inactivated by incubation at 72 °C for 15 min.

Sequencing. Sequencing reactions were performed on a standard high-throughput sequencing system using Big Dye Terminator chemistry (Applied Biosystems) with 2 µl of template cDNA. Each amplicon was sequenced from each end using M13 primers (F primer: TGTAACGACGGCCAGT; R primer: CAGGAAACAGCTATGACC). Sequencing reactions were analysed on an Applied Biosystems 3730 ABI sequencer. Each influenza isolate was processed on its own 96-well plate to minimize the possibility of sample mix-ups.

Data release. Raw traces were submitted to the NCBI Trace Archive. The finished assembly of each isolate, showing how the traces are aligned to one another and to the finished sequence, was deposited in the NCBI Assembly Archive⁶, which allows scientists to investigate the data supporting every nucleotide of each genome. An annotation pipeline developed at NCBI (see Supplementary Methods) was run to make gene assignments, and finished genomes with annotation were deposited without delay in GenBank.

Received 7 July; accepted 16 September 2005.

Published online 5 October 2005.

1. Simonsen, L., Fukuda, K., Schonberger, L. B. & Cox, N. J. The impact of influenza epidemics on hospitalizations. *J. Infect. Dis.* **181**, 831–837 (2000).
2. Thompson, W. W. *et al.* Influenza-associated hospitalizations in the United States. *J. Am. Med. Assoc.* **292**, 1333–1340 (2004).
3. Cox, N. J. & Subbarao, K. Global epidemiology of influenza: past and present. *Annu. Rev. Med.* **51**, 407–421 (2000).
4. Fauci, A. S. Race against time. *Nature* **435**, 423–424 (2005).
5. Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**, 186–194 (1998).
6. Salzberg, S. L., Church, D., DiCuccio, M., Yaschenko, E. & Ostell, J. The genome Assembly Archive: a new public resource. *PLoS Biol.* **2**, E285 (2004).
7. Pop, M., Phillippy, A., Delcher, A. L. & Salzberg, S. L. Comparative genome assembly. *Brief. Bioinform.* **5**, 237–248 (2004).
8. Gajer, P., Schatz, M. & Salzberg, S. L. Automated correction of genome sequence errors. *Nucleic Acids Res.* **32**, 562–569 (2004).
9. Harris, M. E. *et al.* Among 46 near full length HIV type 1 genome sequences from Rakai District, Uganda, subtype D and AD recombinants predominate. *AIDS Res. Hum. Retroviruses* **18**, 1281–1290 (2002).
10. Mikhail, M. *et al.* Full-length HIV type 1 genome analysis showing evidence for HIV type 1 transmission from a nonprogressor to two recipients who progressed to AIDS. *AIDS Res. Hum. Retroviruses* **21**, 575–579 (2005).
11. Holmes, E. C. *et al.* Whole-genome analysis of human influenza A virus reveals multiple persistent lineages and reassortment among recent H3N2 viruses. *PLoS Biol.* **3**, e300 (2005).
12. Jin, H. *et al.* Two residues in the hemagglutinin of A/Fujian/411/02-like influenza viruses are responsible for antigenic drift from A/Panama/2007/99. *Virology* **336**, 113–119 (2005).
13. Lu, B., Zhou, H., Ye, D., Kembler, G. & Jin, H. Improvement of influenza A/Fujian/411/02 (H3N2) virus growth in embryonated chicken eggs by balancing the hemagglutinin and neuraminidase activities, using reverse genetics. *J. Virol.* **79**, 6763–6771 (2005).
14. Colman, P. M., Varghese, J. N. & Laver, W. G. Structure of the catalytic and antigenic sites in influenza virus neuraminidase. *Nature* **303**, 41–44 (1983).
15. Fanning, T. G., Reid, A. H. & Taubenberger, J. K. Influenza A virus neuraminidase: regions of the protein potentially involved in virus-host interactions. *Virology* **276**, 417–423 (2000).
16. Air, G. M., Els, M. C., Brown, L. E., Laver, W. G. & Webster, R. G. Location of antigenic sites on the three-dimensional structure of the influenza N2 virus neuraminidase. *Virology* **145**, 237–248 (1985).
17. Malide, D., Yewdell, J. W., Bennink, J. R. & Cushman, S. W. The export of major histocompatibility complex class I molecules from the endoplasmic reticulum of rat brown adipose cells is acutely stimulated by insulin. *Mol. Biol. Cell* **12**, 101–114 (2001).
18. Gibbs, J. S., Malide, D., Hornung, F., Bennink, J. R. & Yewdell, J. W. The influenza A virus PB1-F2 protein targets the inner mitochondrial membrane via a predicted basic amphipathic helix that disrupts mitochondrial function. *J. Virol.* **77**, 7214–7224 (2003).
19. Chanturiya, A. N. *et al.* PB1-F2, an influenza A virus-encoded proapoptotic mitochondrial protein, creates variably sized pores in planar lipid membranes. *J. Virol.* **78**, 6304–6312 (2004).
20. Marozin, S. *et al.* Antigenic and genetic diversity among swine influenza A H1N1 and H1N2 viruses in Europe. *J. Gen. Virol.* **83**, 735–745 (2002).
21. Xu, X. *et al.* Reassortment and evolution of current human influenza A and B viruses. *Virus Res.* **103**, 55–60 (2004).
22. Lindstrom, S. E., Cox, N. J. & Klimov, A. Genetic analysis of human H2N2 and early H3N2 influenza viruses, 1957–1972: evidence for genetic divergence and multiple reassortment events. *Virology* **328**, 101–119 (2004).
23. Reid, A. H., Taubenberger, J. K. & Fanning, T. G. Evidence of an absence: the genetic origins of the 1918 pandemic influenza virus. *Nature Rev. Microbiol.* **2**, 909–914 (2004).
24. Cyranoski, D. Vaccine sought as bird flu infects humans. *Nature* **422**, 6 (2003).
25. Abbott, A. & Pearson, H. Fear of human pandemic grows as bird flu sweeps through Asia. *Nature* **427**, 472–473 (2004).
26. Li, K. S. *et al.* Genesis of a highly pathogenic and potentially pandemic H5N1 influenza virus in eastern Asia. *Nature* **430**, 209–213 (2004).
27. Szymkowiak, C. *et al.* Rapid method for the characterization of 3' and 5' UTRs of influenza viruses. *J. Virol. Methods* **107**, 15–20 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors acknowledge the laboratory and bioinformatics assistance provided by D. Kosack, C. Hauser, L. Groveman, R. Halpin, A. Phillippy and J. Sparenborg at TIGR, and S. Griesemer, M. Kleabonas and R. Bennett at the Wadsworth Center. We also thank M. Giovanni for help with project oversight and direction. This work was supported in part by the US National Institute of Allergy and Infectious Diseases. Viruses described in this study include some isolates collected as part of the Sentinel Physician Influenza Surveillance Program, which is supported by the US Centers for Disease Control and Prevention.

Author Information GenBank accession numbers are provided in the Supplementary Data. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.L.S. (salzberg@umd.edu).

Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus

Etienne Meylan¹, Joseph Curran², Kay Hofmann³, Darius Moradpour⁴, Marco Binder⁵, Ralf Bartenschlager⁵ & Jürg Tschopp¹

Antiviral immunity against a pathogen is mounted upon recognition by the host of virally associated structures. One of these viral 'signatures', double-stranded (ds) RNA, is a replication product of most viruses within infected cells and is sensed by Toll-like receptor 3 (TLR3) and the recently identified cytosolic RNA helicases RIG-I (retinoic acid inducible gene I, also known as Ddx58) and Mda5 (melanoma differentiation-associated gene 5, also known as Ifih1 or Helicard)¹. Both helicases detect dsRNA, and through their protein-interacting CARD domains, relay an undefined signal resulting in the activation of the transcription factors interferon regulatory factor 3 (IRF3) and NF- κ B. Here we describe Cardif, a new CARD-containing adaptor protein that interacts with RIG-I and recruits IKK α , IKK β and IKK ϵ kinases by means of its C-terminal region, leading to the activation of NF- κ B and IRF3. Overexpression of Cardif results in interferon- β and NF- κ B promoter activation, and knockdown of Cardif by short interfering RNA inhibits RIG-I-dependent antiviral responses. Cardif is targeted and inactivated by NS3-4A, a serine protease from hepatitis C virus known to block interferon- β production. Cardif thus functions as an adaptor, linking the cytoplasmic dsRNA receptor RIG-I to the initiation of antiviral programmes.

Innate immune responses defend the host from invading pathogens by recognizing pathogen-associated molecular patterns. Toll-like receptors are transmembrane receptors that use their leucine-rich repeats to recognize pathogen-associated molecular patterns and initiate intracellular responses through recruitment of TIR-domain containing adaptors (see ref. 2 for a review). TLR3 acts as an antiviral receptor that is capable of recognizing dsRNA. TLR3 transmits cellular responses by specifically recruiting the TIR-containing adaptor protein Trif (also known as TICAM1). Triggering of TLR3 results in a massive type-I interferon (IFN) response, as Trif is able to recruit the essential IRF-3-activating kinases TBK1 and IKK ϵ ^{3,4}. However, the importance of TLR3 in the generation of antiviral immune responses has been recently challenged, as TLR3 is not required to mount antiviral responses against some viruses⁵, suggesting the existence of TLR3-independent dsRNA recognition systems.

Recently, two DexD/H box RNA helicases, RIG-I and Helicard (Mda5), were found to participate in the cytoplasmic recognition of dsRNA^{6,7}. Expression of RIG-I and Helicard can be induced by retinoic acid and IFN- β treatment, respectively. Both helicases contain two amino-terminal caspase recruitment domains (CARD) and a C-terminal region that is characterized by the presence of a helicase domain^{8,9}. RIG-I (and probably Helicard) is able to interact with cytoplasmic dsRNA^{7,10}, and we have previously shown that Helicard is cleaved during apoptosis¹¹. RIG-I and Helicard initiate

antiviral responses by activating the transcription factors IRF3 and NF- κ B, resulting in an enhanced transcription of type-I IFN^{6,7}. The region responsible for RIG-I-induced downstream signalling was mapped to the N-terminal CARD domains, suggesting the existence of a putative CARD adaptor molecule^{7,10} that relays the signal.

In an attempt to identify new CARD-domain-containing proteins, we performed profile searches in human protein databases, which led us to the identification of a protein that we named CARD adaptor inducing IFN- β (Cardif). Cardif contains 540 amino acids and has a N-terminal CARD domain similar to the CARD domains of RIG-I and Helicard (Fig. 1a, b). Cardif corresponds to a gene previously identified in a complementary DNA library screen of potential NF- κ B-activating molecules¹². Polymerase chain reaction with reverse transcription (RT-PCR) of different human cell lines, primary cells and tissues revealed broad expression of Cardif (Supplementary Fig. 1).

Because Cardif is characterized by the presence of a N-terminal CARD domain that is similar to RIG-I and Helicard, we investigated whether the helicases could interact with Cardif upon overexpression of Cardif in human embryonic kidney (HEK) 293T cells. Cardif interacted with RIG-I and showed a weak interaction with Helicard, but did not interact with any of the other tested CARD-domain-containing proteins (Fig. 1c). Cardif oligomerization through its CARD domain was not detected under these experimental conditions.

In subsequent experiments we concentrated on the functional consequences of the RIG-I–Cardif interaction. Overexpression of full-length Cardif in 293T cells resulted in dose-dependent activation of NF- κ B- and IFN- β -reporter plasmids, as well as the IFN-stimulated response element (ISRE) of ISG54, an IRF3-dependent promoter (Fig. 2a). Similar activation was also observed in HeLa cells (data not shown). Overexpression of a deletion construct containing only the CARD domain did not result in activation of any of the reporter genes tested. However, a construct lacking the N-terminal CARD domain (Δ CARD) was sufficient to drive activation of the reporter genes, but to a much lesser extent than the wild-type construct. Activation of the ISRE reporter gene was dependent on IRF3 (Supplementary Fig. 2), as a dominant-negative version of IRF3 lacking the N-terminal DNA-binding region (IRF3 Δ N) inhibited Cardif- and also Trif-induced activation of ISRE, but did not prevent NF- κ B activation. A non-phosphorylatable form of I κ B (dominant-negative I κ B) inhibited Cardif- and Trif-induced NF- κ B activation only. Overexpression of Cardif was sufficient to induce transcription of interferon- β (*Ifnb1*) and of the NF- κ B target gene interleukin 6 (*Il6*) in mouse embryonic fibroblasts (MEFs), as well as inducing the activation of endogenous IRF3 by phosphorylation¹³ and

¹Department of Biochemistry, University of Lausanne, BIL Biomedical Research Center, Chemin des Boveresses 155, CH-1066 Epalinges, Switzerland. ²Department of Microbiology and Molecular Medicine, Faculty of Medicine, University of Geneva, 1 rue Michel Servet, CH-1211 Geneva 4, Switzerland. ³MEMOREC Biotec GmbH, a Miltenyl Biotec company, Stoeckheimer Weg 1, D-50829 Koeln, Germany. ⁴Division of Gastroenterology and Hepatology, Centre Hospitalier Universitaire Vaudois, Rue du Bugnon 46, CH-1011 Lausanne, Switzerland. ⁵Department of Molecular Virology, University Heidelberg, Im Neuenheimer Feld 345, 69120 Heidelberg, Germany.

IRF3-dependent production of RANTES in 293T cells (Fig. 2b, c).

Together, these results suggest that the carboxy-terminal region of Cardif might be implicated in the recruitment and activation of IKK kinase complexes that are essential for the activation of NF- κ B and IRF3. Indeed, we found that Cardif interacted with IKK ϵ (which is able to phosphorylate IRF3) and with IKK α and IKK β (which are known to phosphorylate I κ B), but not with TBK1 (Fig. 3a). This recruitment was more evident for the Cardif mutant containing a deleted CARD domain than for the full-length protein. The Cardif-IKK ϵ interaction contrasts with the propensity of Trif to preferentially interact with TBK1 as opposed to IKK ϵ ¹⁴. This suggests that Trif and Cardif use two different IKK kinases to phosphorylate IRF3, although this might be cell-type-specific.

To unambiguously demonstrate that Cardif acts downstream of RIG-I in antiviral signalling, we overexpressed RIG-I in 293T cells to

induce activation of the IFN- β luciferase promoter, and monitored the consequence of using siRNA to knock down the expression of Cardif, RIG-I and Trif (Fig. 3b). Western blotting showed that all three siRNAs efficiently reduced the protein levels of their respective targets (Supplementary Fig. 3). As a consequence of Cardif knock-down, RIG-I-induced, but not Trif-induced IFN- β promoter activation was greatly reduced (Fig. 3b). Notably, although Trif siRNA inhibited Trif-induced IFN- β promoter activation, it increased the IFN- β response triggered by the overexpression of RIG-I, suggesting that RIG-I may act as a sensor of siRNA within the cytoplasm.

Next, we sought to determine whether Cardif siRNA can inhibit an IFN response in the context of stimulation with dsRNA or infection with Sendai virus (SenV), a virus that triggers RIG-I-dependent IFN responses¹⁰. As neither Poly(I:C) transfection nor SenV infection of 293T cells resulted in an activation of the IFN- β promoter *per se*, we

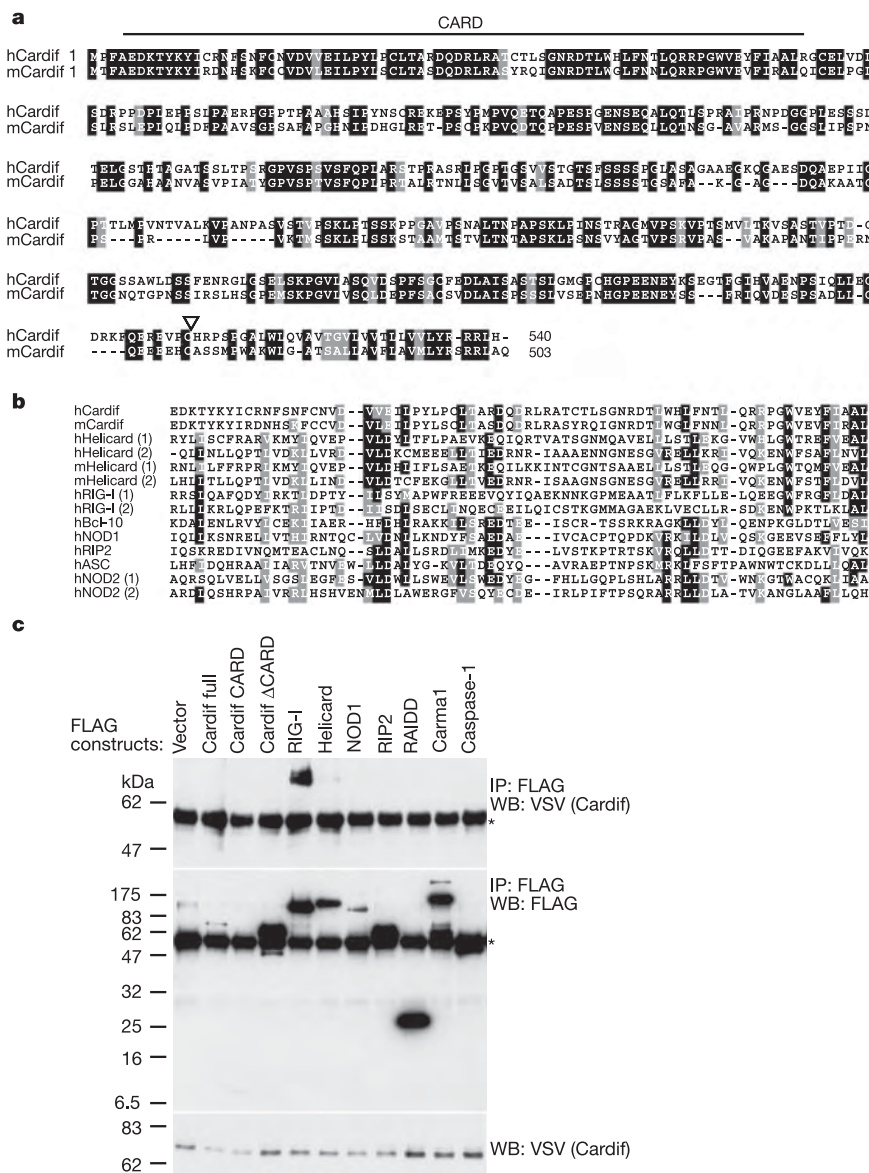


Figure 1 | Cardif is a novel CARD protein that interacts with RIG-I.

a, Amino acid sequence alignment of human (h) and murine (m) Cardif. Shading indicates 100% amino acid sequence identity (black) or similarity (grey). The N-terminal CARD domain is indicated. The cleavage site targeted by NS3-4A is indicated with an arrowhead. **b**, Sequence alignment of the CARD domains of human and murine Cardif and Helicard (Mda5), and human RIG-I, Bcl-10, NOD1, RIP2, ASC and NOD2. Numbers in parentheses indicate the CARD domain number. Shading indicates $\geq 50\%$

amino acid sequence identity (black) and similarity (grey). The CARD domain of Cardif shares 37% and 29% similarity with the two CARD domains of Helicard, and 32% and 29% similarity with the CARD domains of RIG-I. **c**, HEK 293T cells were transfected with the indicated FLAG-tagged constructs or an empty plasmid (vector), together with VSV-Cardif. Anti-FLAG immunoprecipitates (IP) and cell extracts were then analysed by western blot (WB). Asterisk indicates immunoglobulin heavy chain.

reasoned that RIG-I, which is an inducible gene, might not be expressed under these conditions. We therefore transfected a small amount of RIG-I, which permitted us to observe a synergistic effect of SenV or Poly(I:C) on IFN activation (Fig. 3c, d). Cardif knockdown inhibited both Poly(I:C)- and SenV-induced IFN- β promoter activation. To confirm these findings, vesicular stomatitis virus (VSV) was used to infect primary MEFs that depend on RIG-I to mount an antiviral response¹⁵. Cardif knockdown by siRNA inhibited VSV-triggered RANTES expression (Fig. 3e). Together, these results demonstrate the important role of Cardif as an adaptor of RIG-I-dependent antiviral immune responses.

Several human pathogenic viruses, including influenza virus, vaccinia virus and Ebola virus, have evolved strategies to inhibit the early signalling events that lead to IFN production¹⁶. NS3-4A, a multifunctional protein of hepatitis C virus (HCV) that has serine protease activity essential for the production of mature viral proteins, has recently been shown to block the activation of IRF3 (ref. 17). NS3-4A counteracts TLR3-dependent pathways by targeting Trif for proteolytic cleavage, and also counteracts TLR3-independent pathways by targeting an undefined protein of the RIG-I-dependent pathway^{18,19}. IKK ϵ overexpression was shown to overcome NS3-4A-mediated block of IRF3 activation, suggesting that this viral protease targets an adaptor protein 'upstream' of IKK ϵ ²⁰.

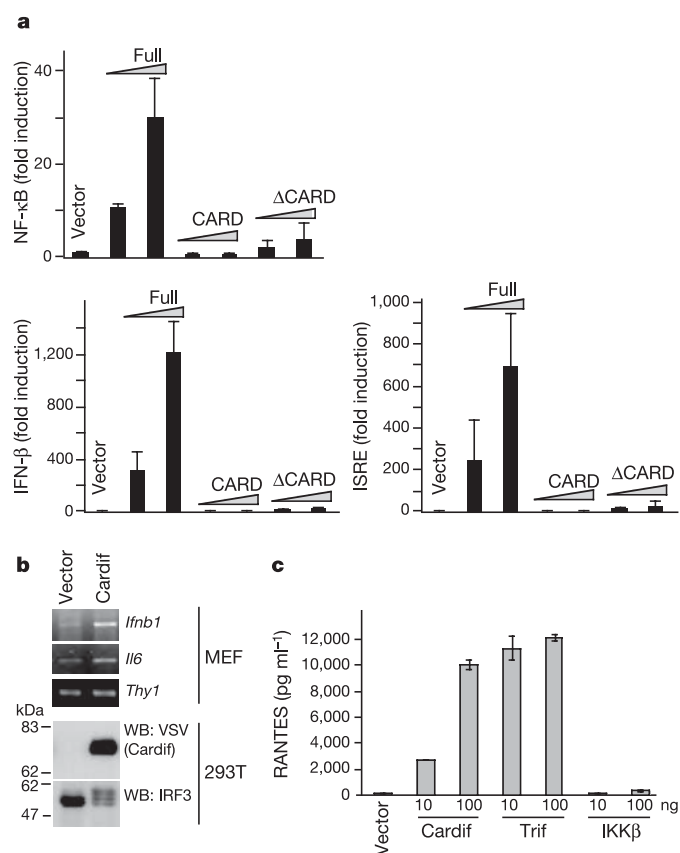


Figure 2 | Cardif activates NF- κ B and IRF3. **a**, 293T cells were transfected with NF- κ B, IFN- β or ISRE reporter plasmids, together with an empty vector or increasing concentrations of the indicated Cardif constructs (full-length, CARD or Δ CARD), and analysed for NF- κ B-, IFN- β - and ISRE-dependent luciferase activity. **b**, MEFs (top) or 293T cells (bottom) were transfected with an empty vector or VSV-Cardif. Total RNA was isolated from MEFs and assessed for the expression of *Ifnb1*, interleukin 6 (*Il6*) and *Thy1* by RT-PCR. For 293T cells, cell extracts were analysed by western blot. **c**, 293T cells were transfected with an empty vector, or with 10 or 100 ng of Cardif, Trif or IKK β cDNA. The supernatants were analysed by ELISA for endogenous expression of human RANTES. Error bars in **a** and **c** represent s.d. of triplicate determinations in a single experiment.

We therefore hypothesized that Cardif might be a substrate for NS3-4A. Indeed, Cardif is cleaved by wild-type but not by a catalytically inactive form of NS3-4A (Fig. 4a). Trif was also cleaved by NS3-4A under the same conditions, as previously reported¹⁸. The N- and C-terminally tagged cleavage fragments of Cardif allowed us to map the cleavage site to within 5 kDa of the C terminus (Fig. 4a).

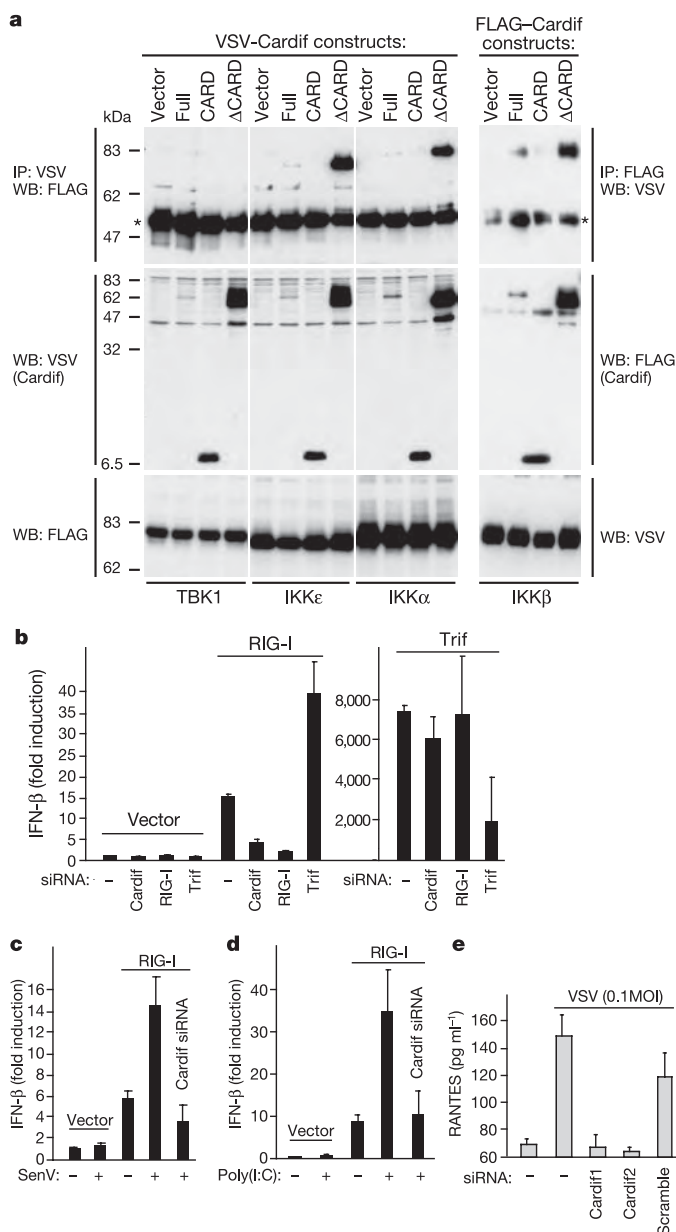


Figure 3 | Cardif links RIG-I to antiviral responses. **a**, 293T cells were transfected with FLAG-TBK1, FLAG-IKK α , FLAG-IKK ϵ and VSV-Cardif constructs (full-length, CARD or Δ CARD), or VSV-IKK β and FLAG-Cardif constructs or an empty vector. Tagged Cardif was immunoprecipitated and cell extracts were analysed by western blot. Asterisk indicates immunoglobulin heavy chain. **b–d**, 293T cells were analysed for IFN- β -dependent luciferase activity after co-transfection with (1) an IFN- β reporter plasmid, (2) the indicated siRNAs, and (3) RIG-I, Trif or an empty vector (**b**), or after co-transfection as above and infection with Sendai virus (SenV) (**c**), or after co-transfection as above and also with Poly(I:C) (**d**). **e**, Primary MEFs were transfected with two different siRNAs targeting murine Cardif, or a control siRNA (scramble), and were infected 30 h later with VSV for 18 h. The supernatants were analysed by ELISA for endogenous expression of murine RANTES. Error bars in **b–e** represent s.d. of triplicate determinations in a single experiment.

As NS3-4A is known to specifically cleave targets after Cys or Thr residues²¹, we mutated the most likely target cysteine, located in the C-terminal region of Cardif (Cys 508, see Fig. 1a), to alanine (C508A). This construct showed complete resistance to cleavage by NS3-4A (Fig. 4a, right). Wild-type Cardif, but not the C508A mutant form, was also cleaved when the entire HCV polyprotein was

inducibly expressed in a U-2 OS-derived tetracycline-regulated cell line (Fig. 4b). Using a recently described *in vitro* HCV infection system^{22–24}, we also observed Cardif cleavage in cells harbouring infectious HCV (Fig. 4c). Upon overexpression in 293T cells, both wild-type and C508A mutant Cardif, but not a Cardif construct lacking the C-terminal cleavage-generated end (Cardif 1–508),

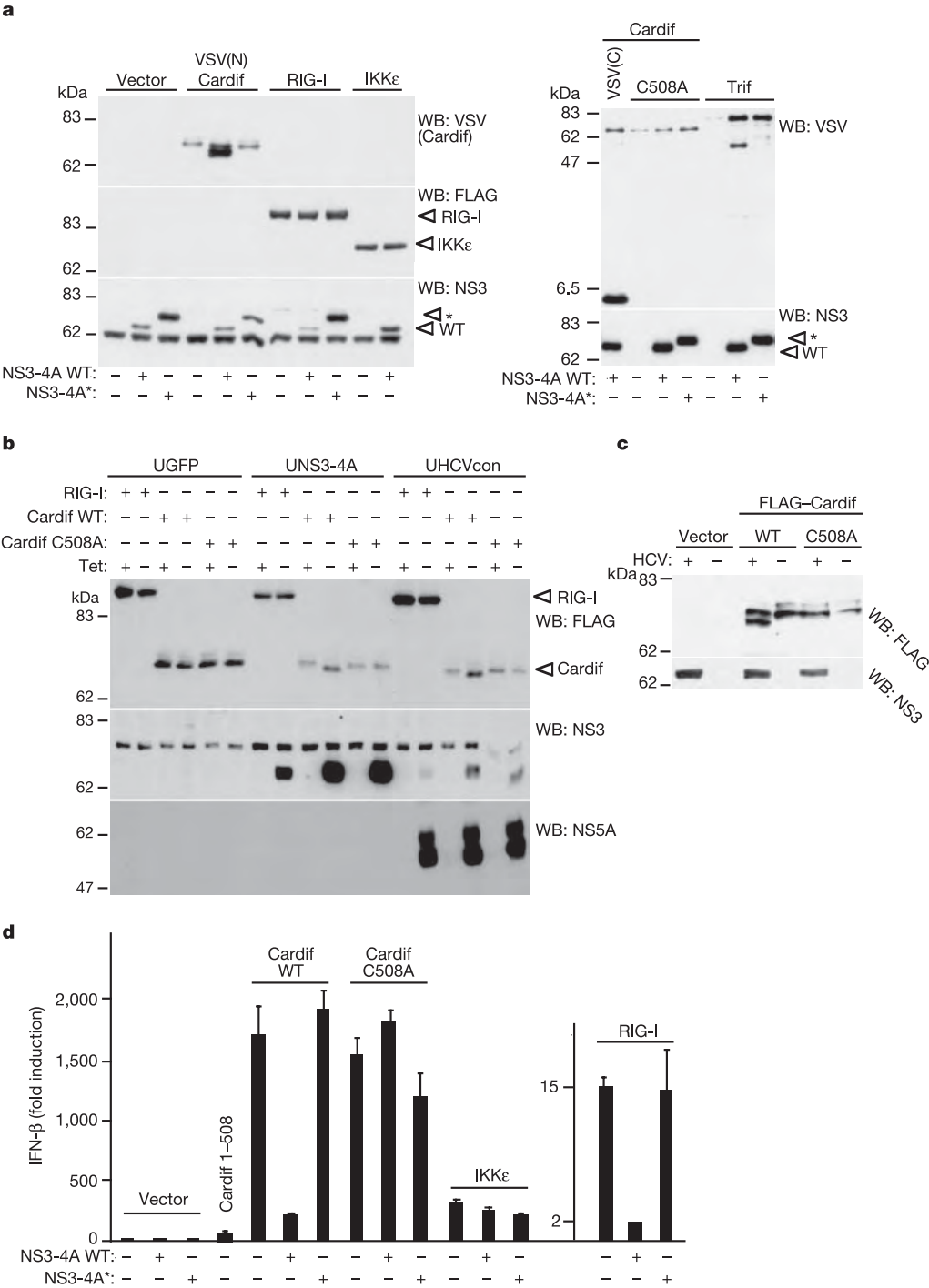


Figure 4 | Cardif is cleaved by the HCV protease NS3-4A. **a**, 293T cells were transfected with FLAG-RIG-I or FLAG-IKKε, VSV-Trif, VSV-Cardif (wild type (WT), either N-terminally VSV(N) or C-terminally VSV(C) tagged, or a mutant Cardif C508A), together with wild-type (WT) or catalytically inactive (*) NS3-4A, and cell extracts were analysed by western blot. **b**, U-2 OS-derived tetracycline-regulated cells were transfected with FLAG-RIG-I and FLAG-Cardif (WT or mutant C508A). After induction of the viral proteins (see Methods), cell extracts were analysed for Cardif cleavage by

western blot. **c**, Huh7.5 cells were transfected with wild-type or C508A mutant Cardif. The cells were subsequently infected with HCV Jc1, and analysed for Cardif cleavage by western blot. **d**, 293T cells were co-transfected with (1) an IFN-β reporter plasmid, (2) an empty vector, C-terminally truncated Cardif (1–508), wild-type Cardif, mutant (C508A) Cardif, IKKε or RIG-I, and (3) the indicated NS3-4A constructs, and analysed for IFN-β-dependent luciferase activity. Error bars represent s.d. of triplicate determinations in a single experiment.

triggered an IFN- β response (Fig. 4d). Catalytically active wild-type NS3-4A protease (NS3-4A) potentially impaired the ability of wild-type Cardif and RIG-I—but not Cardif (C508A) or IKK ϵ —to mediate signalling (Fig. 4d), demonstrating that NS3-4A-mediated cleavage at Cys 508 results in the inactivation of Cardif.

Millions of people are chronically infected with HCV and are at risk of developing liver disease. HCV has evolved an efficient strategy to block a dsRNA-dependent innate immune response, by specifically cleaving and thereby inactivating two adaptor proteins—Trif and Cardif—that have important roles in the two distinct arms of the IFN response mediated by TLR3 and RIG-I, respectively. The potent antiviral activity of BILN 2061, an NS3-4A inhibitor developed for HCV treatment²⁵, was initially explained by its inhibition of the NS3-4A-induced maturation of the HCV polyprotein. In addition, BILN 2061 might prevent Cardif and/or Trif inactivation during infection with HCV, thereby maintaining the innate immune response. Interestingly, a mutation in the Cardif-interacting CARD domain of RIG-I was recently found to allow replication of an HCV replicon in otherwise resistant cells¹⁰, supporting an important role for the RIG-I–Cardif pathway in anti-HCV immunity.

A model consistent with our data includes RIG-I (and Helicard) as a crucial sensor for cytoplasmic dsRNA derived from viruses. Upon binding of dsRNA to RIG-I, a conformational change caused by helicase activity could expose its CARD domain, which associates with the CARD domain of Cardif through a homotypic interaction. As a consequence, the C-terminal segment of Cardif exposes binding sites for the IRF3-phosphorylating kinase IKK ϵ and the NF- κ B-activating kinases IKK α and IKK β , triggering a strong antiviral IFN- β and inflammatory response. Bearing in mind that HCV (and possibly other viruses that are known to inhibit IFN production¹⁶) inactivates the RIG-I–Cardif pathway, identification of therapeutic agents that mimic the CARD-induced conformational change in Cardif may prove a beneficial antiviral strategy.

Note added in proof: Cardif has also recently been identified by two other research groups, and has been named IPS-1 (ref. 31) or MAVS (ref. 32).

METHODS

Expression vectors. The full length Cardif cDNA was amplified from an expressed-sequence-tag (EST) (IMAGE clone 5751684) by standard PCR using Pwo polymerase (Roche). Cardif was subsequently cloned into a derivative of pCR3 (Invitrogen), in-frame with an N- or C-terminal VSV or FLAG tag. Cardif CARD (amino acids 1–95), Δ CARD (amino acids 91–540) and C-terminally truncated (amino acids 1–508) coding sequences were amplified by PCR. Mutation of Cys 508 to Ala (C508A) in Cardif was achieved by site-directed mutagenesis with two sequential rounds of PCR. Mouse *Irf3* lacking the N-terminal DNA-binding domain (Δ N, amino acids 137–419) was amplified from an EST (IMAGE clone 3666172). The fidelity of all PCR amplifications was confirmed by sequencing. FLAG–IKK α , FLAG–IKK ϵ and FLAG–TBK1 were gifts from G. P. Dotto, T. Maniatis, and M. Nakanishi, respectively. FLAG–RIG-I cDNA was provided by T. Fujita. Human TRIF and mouse Helicard (Mda5) expression plasmids have been described^{11,26}. Wild-type pCMVNS3-4A (NS3-4A WT) has been described²⁷, and its catalytically inactive counterpart (NS3-4A*) containing a Ser 139 to Ala mutation will be described elsewhere. NF- κ B-Luc and IFN- β -Luc reporter plasmids were provided by V. Jongeneel and T. Taniguchi, respectively. pISRE-Luc was from Stratagene. The Renilla-luciferase transfection efficiency vector (pRLTK) was purchased from Promega.

PCR with reverse transcription. RT-PCR was performed as described²⁶. Human multiple tissue cDNAs (636742, 636743) were from BD Biosciences. Human primary haematopoietic cells and corresponding cDNAs have been described²⁸. The following primers were used: Cardif, 5'-ACCTTCATTGCGG CACTGAGG-3' and 5'-TCTGGATTCTTGGGATGGC-3'; β -actin, 5'-GGCATCGTGATGGACTCCG-3' and 5'-GCTGGAAGGTGGACAGCGA-3'; mouse *Thy1* 5'-CCATCCAGCATGAGTTCAGCC-3' and 5'-GCATCCAGGAT GTGTCTGA-3'; mouse *Irfn1* 5'-TTCCTGCTGTCTTCTCCAC-3' and 5'-GATTCATACCAGTCCAGAGTC-3'; mouse *Il6* 5'-ATGAAGTCTCTCT GCAAGAGACT-3' and 5'-CACTAGGTTTGGCAGTAGATCTC-3'.

Cell culture conditions. The human embryonic kidney (HEK) 293T cell line, primary mouse embryonic fibroblasts (MEFs) and the hepatoma cell line Huh7.5 (a gift from C.M. Rice) were grown in Dulbecco's modified Eagle's

medium (Invitrogen) supplemented with 10% heat-inactivated fetal calf serum. U-2 OS human osteosarcoma-derived cell lines that inducibly express green fluorescent protein (UGFP) or NS3-4A upon tetracycline withdrawal, either alone (UNS3-4A) or in the context of the entire HCV polyprotein (UHCVcon), have been described^{27,29}.

Transfection, immunoprecipitation and immunoblotting. Transfection, immunoprecipitation and immunoblotting assays in 293T cells were performed as previously described²⁶. siRNAs (Ambion) and Poly(I:C) (Invivogen) were also transfected into 293T cells using the calcium-phosphate precipitation technique, at final concentrations of 10 nM (siRNA) or 1 ng ml⁻¹ (Poly(I:C)). MEFs were transfected using Lipofectamine 2000 (Invitrogen) or, in the case of siRNA, using TransIT-TKO (Mirus), according to the manufacturers' instructions. U-2 OS cells were transfected using the calcium-phosphate precipitation technique. Twenty-four hours after transfection, U-2 OS cells were washed twice with PBS and were then cultured for 24 h in the presence (tet⁺) or absence (tet⁻) of tetracycline to allow the expression of GFP (UGFP), NS3-4A alone (UNS3-4A) or the entire HCV polyprotein (UHCVcon). Huh7.5 cells were transfected using Effectene (Qiagen), according to the manufacturer's instructions. Anti-FLAG M2 and anti-VSV P5D4 antibodies were from Sigma. Anti-IRF3 (sc-9082) was from Santa Cruz. Anti-NS3 1B6 and anti-NS5A 11H have been previously described^{27,30}. For the HCV infection experiment, a polyclonal anti-NS3 antiserum was used.

Viral infections. Twenty-four hours after transfection, 293T cells were infected for 24 h with Sendai virus (strain H) at a multiplicity of infection (MOI) of 10. Cells were subsequently lysed and assayed for luciferase activity. Thirty hours after transfection with siRNA, primary MEFs were infected for 18 h with VSV at a MOI of 0.1. HCV chimaeric genotype 2a virus, designated Jc1, will be described elsewhere. HCV production was performed as previously described²². Six hours after transfection, Huh7.5 cells were infected for 5 h with cell culture supernatants containing infectious Jc1 particles, cultured for another 60 h and harvested in Laemmli sample buffer for analysis by western blot.

Luciferase reporter assays. Luciferase reporter assays were performed as described previously²⁶.

Analysis of RANTES expression. Forty-eight hours after transfection of 293T cells in 96-well plates, cell supernatants were analysed for human RANTES expression by enzyme-linked immunosorbent assay (ELISA, R&D Systems), according to the manufacturer's instructions. Eighteen hours after infection of MEFs with VSV in 12-well plates, cell supernatants were analysed for murine RANTES expression (murine RANTES ELISA, R&D Systems), according to the manufacturer's instructions.

Received 25 August; accepted 6 September 2005.

Published online 21 September 2005.

1. Levy, D. E. & Marie, I. J. RIGging an antiviral defense—it's in the CARDs. *Nature Immunol.* **5**, 699–701 (2004).
2. Akira, S. & Takeda, K. Toll-like receptor signalling. *Nature Rev. Immunol.* **4**, 499–511 (2004).
3. Sharma, S. *et al.* Triggering the interferon antiviral response through an IKK-related pathway. *Science* **300**, 1148–1151 (2003).
4. Fitzgerald, K. A. *et al.* IKK ϵ and TBK1 are essential components of the IRF3 signalling pathway. *Nature Immunol.* **4**, 491–496 (2003).
5. Edelmann, K. H. *et al.* Does Toll-like receptor 3 play a biological role in virus infections? *Virology* **322**, 231–238 (2004).
6. Andrejeva, J. *et al.* The V proteins of paramyxoviruses bind the IFN-inducible RNA helicase, mda-5, and inhibit its activation of the IFN- β promoter. *Proc. Natl Acad. Sci. USA* **101**, 17264–17269 (2004).
7. Yoneyama, M. *et al.* The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nature Immunol.* **5**, 730–737 (2004).
8. Sun, Y. W. *RIG-I, a Human Homolog Gene of RNA Helicase, is Induced by Retinoic Acid During the Differentiation of Acute Promyelocytic Leukemia Cell*. Thesis, Shanghai Institute of Hematology, Shanghai Second Medical University (1997).
9. Kang, D. C. *et al.* mda-5: An interferon-inducible putative RNA helicase with double-stranded RNA-dependent ATPase activity and melanoma growth-suppressive properties. *Proc. Natl Acad. Sci. USA* **99**, 637–642 (2002).
10. Sumpter, R. Jr *et al.* Regulating intracellular antiviral defense and permissiveness to hepatitis C virus RNA replication through a cellular RNA helicase, RIG-I. *J. Virol.* **79**, 2689–2699 (2005).
11. Kovacsovics, M. *et al.* Overexpression of Helicard, a CARD-containing helicase cleaved during apoptosis, accelerates DNA degradation. *Curr. Biol.* **12**, 838–843 (2002).
12. Matsuda, A. *et al.* Large-scale identification and characterization of human genes that activate NF- κ B and MAPK signalling pathways. *Oncogene* **22**, 3307–3318 (2003).
13. Servant, M. J. *et al.* Identification of the minimal phosphoacceptor site required for *in vivo* activation of interferon regulatory factor 3 in response to virus and double-stranded RNA. *J. Biol. Chem.* **278**, 9441–9447 (2003).

14. Sato, S. *et al.* Toll/IL-1 receptor domain-containing adaptor inducing IFN- β (TRIF) associates with TNF receptor-associated factor 6 and TANK-binding kinase 1, and activates two distinct transcription factors, NF- κ B and IFN-regulatory factor-3, in the Toll-like receptor signalling. *J. Immunol.* **171**, 4304–4310 (2003).
15. Kato, H. *et al.* Cell type-specific involvement of RIG-I in antiviral response. *Immunity* **23**, 19–28 (2005).
16. Conzelmann, K. K. Transcriptional activation of alpha/beta interferon genes: interference by nonsegmented negative-strand RNA viruses. *J. Virol.* **79**, 5241–5248 (2005).
17. Foy, E. *et al.* Regulation of interferon regulatory factor-3 by the hepatitis C virus serine protease. *Science* **300**, 1145–1148 (2003).
18. Li, K. *et al.* Immune evasion by hepatitis C virus NS3/4A protease-mediated cleavage of the Toll-like receptor 3 adaptor protein TRIF. *Proc. Natl Acad. Sci. USA* **102**, 2992–2997 (2005).
19. Foy, E. *et al.* Control of antiviral defenses through hepatitis C virus disruption of retinoic acid-inducible gene-I signalling. *Proc. Natl Acad. Sci. USA* **102**, 2986–2991 (2005).
20. Breiman, A. *et al.* Inhibition of RIG-I-dependent signalling to the interferon pathway during hepatitis C virus expression and restoration of signalling by IKKe. *J. Virol.* **79**, 3969–3978 (2005).
21. Bartenschlager, R., Ahlborn-Laake, L., Yasargil, K., Mous, J. & Jacobsen, H. Substrate determinants for cleavage in *cis* and in *trans* by the hepatitis C virus NS3 proteinase. *J. Virol.* **69**, 198–205 (1995).
22. Wakita, T. *et al.* Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nature Med.* **11**, 791–796 (2005).
23. Zhong, J. *et al.* Robust hepatitis C virus infection *in vitro*. *Proc. Natl Acad. Sci. USA* **102**, 9294–9299 (2005).
24. Lindenbach, B. D. *et al.* Complete replication of hepatitis C virus in cell culture. *Science* **309**, 623–626 (2005).
25. Lamarre, D. *et al.* An NS3 protease inhibitor with antiviral effects in humans infected with hepatitis C virus. *Nature* **426**, 186–189 (2003).
26. Meylan, E. *et al.* RIP1 is an essential mediator of Toll-like receptor 3-induced NF- κ B activation. *Nature Immunol.* **5**, 503–507 (2004).
27. Wolk, B. *et al.* Subcellular localization, stability, and *trans*-cleavage competence of the hepatitis C virus NS3–NS4A complex expressed in tetracycline-regulated cell lines. *J. Virol.* **74**, 2293–2304 (2000).
28. Schneider, P. *et al.* BAFF, a novel ligand of the tumour necrosis factor family, stimulates B cell growth. *J. Exp. Med.* **189**, 1747–1756 (1999).
29. Schmidt-Mende, J. *et al.* Determinants for membrane association of the hepatitis C virus RNA-dependent RNA polymerase. *J. Biol. Chem.* **276**, 44052–44063 (2001).
30. Brass, V. *et al.* An amino-terminal amphipathic α -helix mediates membrane association of the hepatitis C virus nonstructural protein 5A. *J. Biol. Chem.* **277**, 8130–8139 (2002).
31. Kawai, T. *et al.* IPS-1, an adaptor triggering RIG-I- and Mda5-mediated type I interferon induction. *Nature Immunol.* advance online publication, 28 August 2005 (doi:10.1038/nl243).
32. Seth, R. B., Sun, L., Ea, C. K. & Chen, Z. J. Identification and characterization of MAVS, a mitochondrial antiviral signalling protein that activates NF- κ B and IRF3. *Cell* **122**, 669–682 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Hertig for technical support, and H. Everett, V. Petrilli and O. Donzé for discussions and critical reading of the manuscript. This work was supported by grants from the Swiss National Science Foundation and the Commission of Technology and Innovation (CTI).

Author Information The DNA sequence of human Cardif has been deposited in GenBank under accession number DQ181928. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.T. (jurg.tschopp@unil.ch).

Towards a proteome-scale map of the human protein-protein interaction network

Jean-François Rual^{1*}, Kavitha Venkatesan^{1*}, Tong Hao¹, Tomoko Hirozane-Kishikawa¹, Amélie Dricot¹, Ning Li¹, Gabriel F. Berriz², Francis D. Gibbons², Matija Dreze^{1,3}, Nono Ayivi-Guedehoussou¹, Niels Klitgord¹, Christophe Simon¹, Mike Boxem¹, Stuart Milstein¹, Jennifer Rosenberg¹, Debra S. Goldberg², Lan V. Zhang², Sharyl L. Wong², Giovanni Franklin², Siming Li^{1†}, Joanna S. Albala^{1†}, Janghoo Lim⁴, Carlene Fraughton¹, Estelle Llamasas¹, Sebiha Cevik¹, Camille Bex¹, Philippe Lamesch^{1,3}, Robert S. Sikorski⁵, Jean Vandenhaute³, Huda Y. Zoghbi⁴, Alex Smolyar¹, Stephanie Bosak⁶, Reynaldo Sequerra⁶, Lynn Doucette-Stamm⁶, Michael E. Cusick¹, David E. Hill¹, Frederick P. Roth² & Marc Vidal¹

Systematic mapping of protein-protein interactions, or 'interactome' mapping, was initiated in model organisms, starting with defined biological processes^{1,2} and then expanding to the scale of the proteome³⁻⁷. Although far from complete, such maps have revealed global topological and dynamic features of interactome networks that relate to known biological properties^{8,9}, suggesting that a human interactome map will provide insight into development and disease mechanisms at a systems level. Here we describe an initial version of a proteome-scale map of human binary protein-protein interactions. Using a stringent, high-throughput yeast two-hybrid system, we tested pairwise interactions among the products of ~8,100 currently available Gateway-cloned open reading frames and detected ~2,800 interactions. This data set, called CCSB-HI1, has a verification rate of ~78% as revealed by an independent co-affinity purification assay, and correlates significantly with other biological attributes. The CCSB-HI1 data set increases by ~70% the set of available binary interactions within the tested space and reveals more than 300 new connections to over 100 disease-associated proteins. This work represents an important step towards a systematic and comprehensive human interactome project.

Our working definition of a human interactome map is the complete collection of binary protein-protein interactions detectable in one or more exogenous assay. This definition excludes dynamic and functional properties of these interactions (Supplementary Data I). Thus, we treat interactome maps as 'scaffold' information, from which increasingly detailed and reliable biological models can be generated by integrating other functional genomic and proteomic data sets¹⁰ (Supplementary Data II).

The currently available information on the human interactome network originates from either literature-curated (LC) interactions¹¹⁻¹⁵, or from 'interologs' (that is, potential interactions predicted from interactome data available for model organisms given evolutionary conservation of two known partners)^{2,16}. This information needs to be complemented by systematic experimental mapping approaches that are: (1) not biased towards any particular biological interest (that is, without 'inspection bias'), as is the case for

LC data sets; (2) more complete; and (3) supported by experiments rather than predictions. We are mapping the human interactome network systematically in successive versions, with each version defined by the availability of recombinationally cloned open reading frames (ORFs) in the human 'ORFeome'¹⁷.

In this initial version, we use human ORFeome v1.1 (ref. 17), a resource containing ~8,100 Gateway-cloned ORFs (generated using the Mammalian Gene Collection, as previously described¹⁷) that correspond to ~7,200 distinct protein-coding genes (Supplementary Table S1). Thus, our initial 'search space' (Space-I) encompasses protein pairs encoded by a 7,200 × 7,200 matrix of genes. Future interactome versions can be generated by successively increasing the search space as additional versions of the human ORFeome become available (Supplementary Data III). Accepting a total of ~22,000 protein-coding genes in the human genome¹⁸ and excluding polymorphic and splice variants, Space-I corresponds to ~10% of the total search space for a comprehensive human interactome map (Fig. 1a). Currently, 4,067 binary LC interactions are available in Space-I (LCI interactions; Supplementary Table S2).

Our high-throughput yeast two-hybrid system is highly specific, benefiting from the following features, which were not uniformly present in earlier large-scale studies: relatively low levels of expression of both Gal4 DNA-binding domain (DB) and Gal4 activation domain (AD) hybrid proteins (DB-X and AD-Y, or DB-ORF and AD-ORF); three different yeast two-hybrid inducible reporter genes; and a plasmid-shuffling counter selection to eliminate systematically *de novo* auto-activators¹⁹ (Supplementary Data IV). We tested each of the ~8,100 individual DB-X proteins against 45 mini-libraries, each containing a pool of 188 AD-Y fusion proteins (AD-188Ys), by yeast-mating in a 96-well format (Fig. 1a). Such small pools offer high sensitivity, because positive clones are less likely to be masked by other AD-Y clones within the same pool. Indeed, our overall reproducibility rate was ~55%, close to that observed in proteome-scale affinity purification followed by mass spectrometry experiments²⁰ (Supplementary Data V).

In our Space-I yeast two-hybrid matrix, we identified ~65,000 primary positive colonies, of which 12,251 scored positive after

¹Center for Cancer Systems Biology and Department of Cancer Biology, Dana-Farber Cancer Institute and Department of Genetics, Harvard Medical School, 44 Binney Street, Boston, Massachusetts 02115, USA. ²Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 250 Longwood Ave, Boston, Massachusetts 02115, USA. ³Unité de Recherche en Biologie Moléculaire, Facultés Notre-Dame de la Paix, 61 Rue de Bruxelles, 5000 Namur, Belgium. ⁴Howard Hughes Medical Institute, and Departments of Pediatrics, Neurology, Neuroscience, and Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA. ⁵Arcbay, Inc., 6 Whittier Place, Suite 7J, Boston, Massachusetts 01915, USA. ⁶Agencourt Bioscience Corporation, 500 Cummings Center, Suite 2450, Beverly, Massachusetts 01915, USA. [†]Present addresses: ArQule, Inc., 19 Presidential Way, Woburn, Massachusetts 01081, USA (S.L.); Departments of Cancer Biology, and Otolaryngology, Head and Neck Surgery, University of California Davis, 2521 Stockton Blvd, Suite 7200, Sacramento, California 95817, USA (J.S.A.).

*These authors contributed equally to this work.

stringent phenotype testing; that is, we only retained clones that were positive for at least two yeast two-hybrid reporter assays and we controlled for auto-activation (Fig. 1a). Both DB-X and AD-Y fragments from these two-reporter-positive colonies were amplified by polymerase chain reaction (PCR) and sequenced to generate 10,597 pairs of interaction sequence tags (ISTs) (Supplementary Fig. S1a). We then collapsed yeast two-hybrid ISTs corresponding to the same pair of genes and removed lower confidence interactions (Supplementary Data VI). This resulted in a data set containing 2,754 yeast two-hybrid interactions—the Center for Cancer Systems Biology Human Interactome version 1 (CCSB-HI1) (Supplementary Fig. S1a and Supplementary Table S2. All data is publicly available (Supplementary Data VII). In all, CCSB-HI1 provides interaction information for 1,549 proteins, ~21% of the proteins tested in Space-1.

To measure the specificity of CCSB-HI1, we considered technical and biological false positives separately (Supplementary Data VIII). Technical false positives arise from experimental errors that can and should be avoided. To estimate our technical false positive rate, representative samples of interactions were verified by *in vivo* co-affinity purification glutathione *S*-transferase (GST) pull-down assay in human 293T cells (Supplementary Data IX). The verification rates for co-affinity purification were ~78% for yeast two-hybrid-only interactions, ~62% for LCI-only interactions and ~81% for interactions present in both LCI and yeast two-hybrid data sets (Fig. 1b; see also Supplementary Fig. S1b, c and Supplementary Tables S2 and S3). Given these results and the fact that co-affinity purification GST pull-down assays are not perfectly sensitive, we argue that CCSB-HI1 is largely free of technical false positives, and comparable in reliability to LCI interactions, supporting the validity of our improvements of

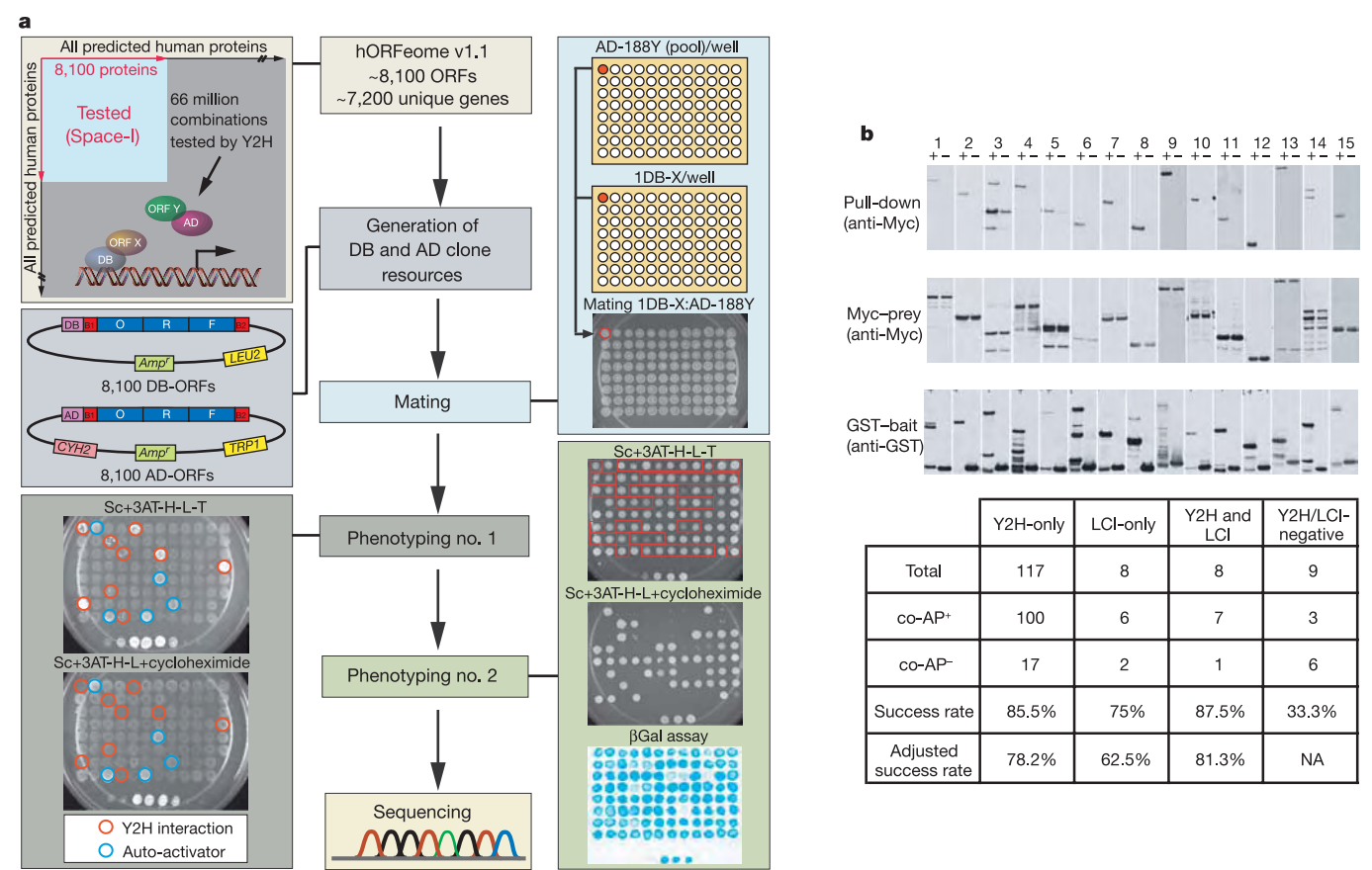


Figure 1 | Towards the generation of a proteome-scale human yeast two-hybrid map. a, Schema of the high-throughput yeast two-hybrid pipeline. Individual steps (middle column) and representative examples (flanking left and right columns) are indicated. The top panel of the left column represents the matrix of all protein pairs. All available ORFs from human ORFeome v1.1 were transferred into both DB and AD vectors by recombinational cloning (middle panel of left column). The top panel of the right column shows the mating process, with each bait mated to individual pools of 188 AD-ORFs. Initial phenotypic testing evaluated growth of diploid cells on selective medium in response to enhanced levels of the *GAL1::HIS3* selective marker (bottom panel of left column). All positive diploids from phenotyping no. 1 (red circles) were subsequently tested for activation of both *GAL1::HIS3* and *GAL1::lacZ* reporter genes. Auto-activators were identified by growth on medium containing cycloheximide (bottom panels of left and right columns). Positive colonies from phenotyping no. 2 (outlined in red) were isolated and used to PCR-amplify both DB-ORF and AD-ORF fragments for sequencing. **b**, Verification of yeast two-hybrid interactions by co-affinity purification assays. Fifteen representative examples of co-affinity purification-positive assays are

shown. The middle and bottom panels show expression controls of Myc-prey and GST-bait fusion proteins, respectively. Each lane pair in the top panels shows presence or absence of Myc-prey fusions after affinity purification, demonstrating binding to GST-bait fusion proteins (+) or to GST alone (–). The Table summarizes the data obtained for four different classes of protein pairs. ‘Y2H and LCI’ describes interactions reported in both the yeast two-hybrid and LCI data sets. ‘Y2H/LCI-negative’ describes pairs of proteins that were not reported to interact either in the yeast two-hybrid or in the LCI data sets. Rows indicate the total number of interactions tested and considered for scoring (Total), the number of interactions not verified by co-affinity purification (co-AP[–]), the number of interactions verified by co-affinity purification (co-AP⁺), the proportion of co-affinity purification-positive interactions (success rate), and the adjusted success rate (which accounts for the observation that one-third of all co-affinity purification experiments yield an apparently positive result without regard to whether or not the protein pair truly interacts; see Supplementary Data IX). Identities, lane positions and scoring of all protein pairs tested by co-affinity purification are provided in Supplementary Tables S2 and S3.

the yeast two-hybrid methodology. Estimating biological false positive interactions, which are genuinely observed in one or more assay but do not occur *in vivo*, is more difficult. We partially addressed this by examining the correlation of CCSB-HI1 data with other biological information (see below).

To measure the sensitivity of CCSB-HI1, we selected two high-confidence subsets from among all 4,067 LCI direct binary interactions. LCI-core contains 624 interactions supported by at least two PubMed entries. LCI-hypercore contains 275 interactions supported by at least two PubMed entries and present in at least two curated databases (Supplementary Table S2). Overall, the fractions of LCI, LCI-core and LCI-hypercore interactions found in CCSB-HI1 are 2.3%, 4.6% and 8.4%, respectively (Fig. 2a). These overlaps are larger than expected by chance ($P < 6 \times 10^{-56}$) and are similar to those found for interactome maps in *Caenorhabditis elegans* and *Drosophila melanogaster*^{7,21}. That the fraction of CCSB-HI1 interactions increases markedly with increasingly confident subsets of LCI suggests that literature-derived interactions are variable in quality and should not necessarily be interpreted as a 'gold standard'. Because Space-I represents ~10% of the human network (without accounting for alternative splice variants), and because we detected ~10% of LCI-hypercore interactions, we conclude that the CCSB-HI1 data set contains ~1% of the human interactome (Supplementary Data X).

We represented the union of all CCSB-HI1 and LCI interactions in a network graph in which nodes are proteins and edges are interactions. The main component of this network contains 2,784 nodes and 6,438 edges (Fig. 2b), and shows interactions largely segregated into two neighbourhoods: one enriched for CCSB-HI1 interactions (red edges) and the other enriched for LCI interactions (blue edges). To explore this hypothesis, we calculated, for each node, the fraction of yeast two-hybrid edges within paths of length 1, 2 and 3 (that is, within '1-hop', '2-hop' and '3-hop' neighbourhoods). The distribution of this fraction (Fig. 2c; see also Supplementary Fig. S2) confirms the evidence-type segregation apparent in Fig. 2b. One explanation for this phenomenon is that different biases exist in the CCSB-HI1 and LCI data sets. For example, certain protein classes (such as those involved in cancer) are studied more extensively than others, resulting in an inherent inspection bias in LC data (Supplementary Table S4). Furthermore, the methodologies used to detect interactions (including yeast two-hybrid) each have different biases for example, under-representation of membrane proteins (Supplementary Data X).

The novelty of CCSB-HI1 interactions was evaluated by systematically searching the PubMed and Google Scholar literature databases for co-occurrence of the corresponding gene symbols. More than 85% of the CCSB-HI1 pairs (as compared with only 25% of

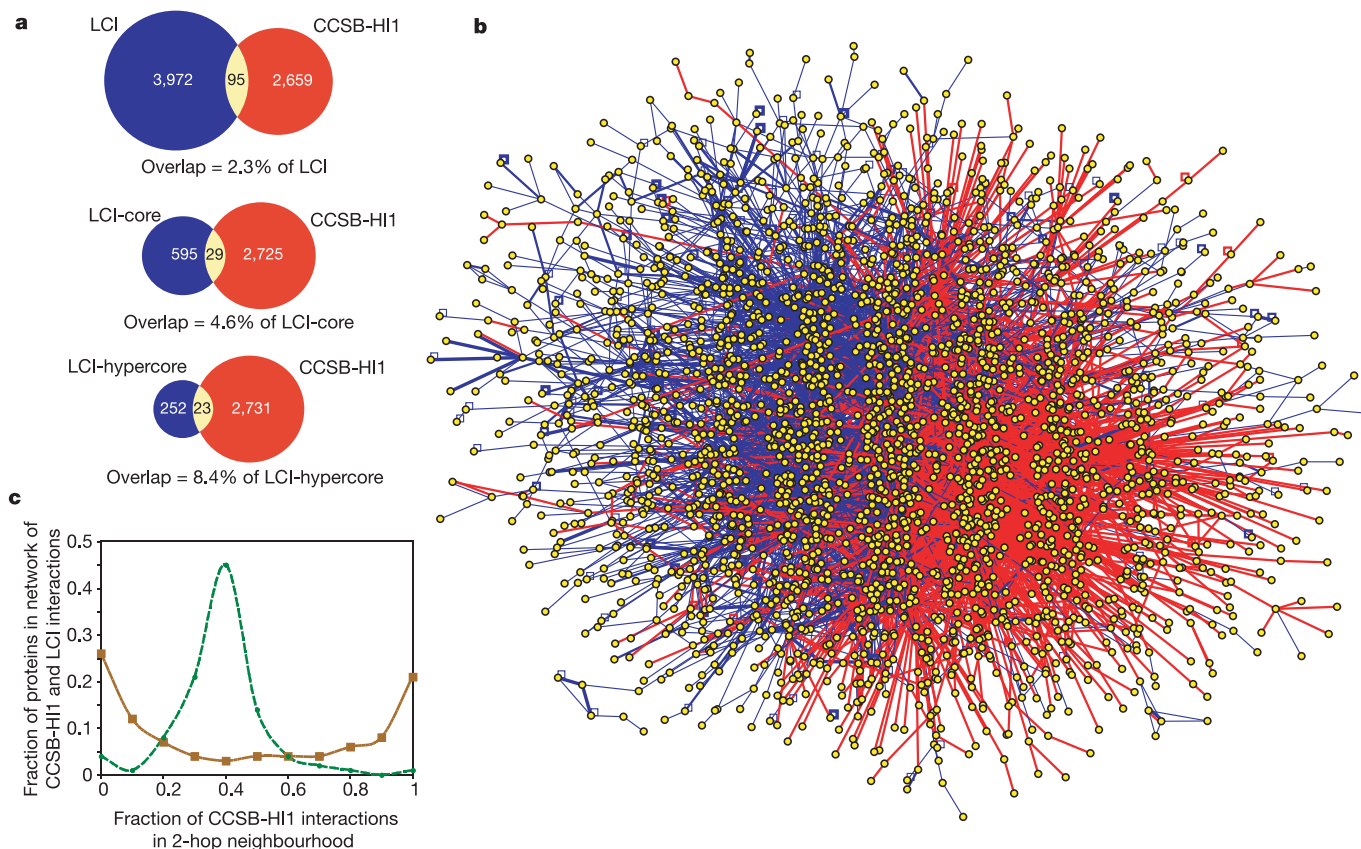


Figure 2 | Overlap of CCSB-HI1 with existing literature-curated (LC) data. a, Overlap between CCSB-HI1 and LC interactions in Space-I (LCI). The top, middle and bottom panels represent the overlap between CCSB-HI1 and LCI, LCI-core and LCI-hypercore, respectively. **b**, Network graph of the union of all CCSB-HI1 and LCI interactions. Proteins are shown as yellow nodes and CCSB-HI1 and LCI interactions are shown as red and blue edges, respectively. Blue edges with increasing thickness indicate LCI-non-core, LCI-core and LCI-hypercore, respectively. The apparent banding pattern of the yellow nodes is an artefact of the graph layout algorithm (Supplementary Data). Importantly, the layout algorithm was not informed by type of supporting evidence and therefore does not explain the

evident separation of blue and red edges. **c**, Bias in 2-hop network neighbourhood for either CCSB-HI1 or LCI interactions. The frequency of nodes with a given proportion of CCSB-HI1 interactions in their 2-hop neighbourhood is depicted for the interactome network graph in **b** (solid curve) and for a network in which the types of supporting evidence (CCSB-HI1 or LCI) are randomly permuted among edges (dashed curve). The solid curve indicates that most of the proteins in the network of **b** have either only CCSB-HI1 or only LCI interactions in their 2-hop neighbourhood. In contrast, neighbourhoods are well mixed when evidence labels are randomly permuted among edges.

Table 1 | Overlap of protein interactions with other gene- or protein-pair characteristics

Protein pairs	Share mouse phenotype		Share upstream motif		Have correlated expression	
	<i>F</i> (C)	<i>P</i> -value	<i>F</i> (C)	<i>P</i> -value	<i>F</i> (C)	<i>P</i> -value
All possible within Space-I	0.128	NA	0.086	NA	0.063	NA
CCSB-HI1	0.257	2.53×10^{-3}	0.115	1.14×10^{-4}	0.130	2.14×10^{-7}
LCI	0.336	4.91×10^{-43}	0.146	9.05×10^{-20}	0.204	5.45×10^{-56}
LCI-core	0.471	7.53×10^{-20}	0.137	3.77×10^{-3}	0.243	3.57×10^{-12}
LCI-non-core	0.306	2.54×10^{-27}	0.147	3.65×10^{-18}	0.198	7.78×10^{-46}

Protein pairs	Share GO component		Share GO function		Share GO process	
	<i>F</i> (C)	<i>P</i> -value	<i>F</i> (C)	<i>P</i> -value	<i>F</i> (C)	<i>P</i> -value
All possible within Space-I	0.059	NA	0.021	NA	0.036	NA
CCSB-HI1	0.488	1.49×10^{-28}	0.250	5.49×10^{-20}	0.233	2.68×10^{-28}
LCI	0.656	6.47×10^{-139}	0.228	5.72×10^{-120}	0.410	1.74×10^{-405}
LCI-core	0.870	5.90×10^{-43}	0.270	1.70×10^{-32}	0.616	3.40×10^{-137}
LCI-non-core	0.610	3.84×10^{-100}	0.218	1.33×10^{-89}	0.368	4.03×10^{-280}

F(C) represents the fraction of gene- or protein-pairs (defined for each row) the given characteristic *C*. Assessed characteristics include shared mouse phenotype, shared upstream motif, correlated expression³⁰ and shared Gene Ontology annotation. ‘All possible within Space-I’ represents all possible gene- or protein-pairs in Space-I for which information regarding *C* is available. For each analysis of a shared characteristic, only gene- or protein-pairs for which both members had some annotation for that characteristic were considered. For analysis of correlated expression, only gene pairs with expression measurements for both genes were considered.



Figure 3 | Interaction network of disease-associated CCSB-HI1 proteins. The network has 121 OMIM disease-associated proteins (green nodes) and 424 CCSB-HI1 interactions involving them (red edges), along with known LC interactions (solid blue edges represent binary LCI interactions and dashed blue edges represent non-binary interactions). Proteins without an OMIM disease association are depicted as yellow nodes, and blue edges with increasing thickness indicate LCI-non-core, LCI-core and LCI-hypercore interactions, respectively. We note that 94 out of the 424 CCSB-HI1 interactions involve the Ewing sarcoma related protein (EWSR1; also known as EWS).

LCI pairs) showed no linkage of the corresponding gene symbols (Supplementary Fig. S3). These results indicate that most of our yeast two-hybrid interactions are novel.

To determine whether messenger RNAs corresponding to interacting protein pairs are likely to be co-expressed, we used Pearson correlation coefficients of the corresponding gene pairs in the CCSB-HI1 and LCI data sets from four expression studies in diverse human and mouse tissues (Supplementary Data XI). LCI pairs were enriched for correlated expression in all four cases ($P < 3 \times 10^{-17}$) and CCSB-HI1 pairs were enriched in three of the four cases ($P < 3 \times 10^{-5}$) (Table 1; see also Supplementary Fig. S4a and Supplementary Tables S2 and S5). In addition, CCSB-HI1 interactions are more enriched than would be expected by chance for (1) presence of a common upstream DNA sequence that is conserved across human, mouse, rat and dog genomes²² ($P = 1 \times 10^{-4}$), (2) orthologous genes in mouse having a specific phenotype in common²³ ($P = 3 \times 10^{-3}$), and (3) annotation with the same Gene Ontology (GO) terms²⁴ ($P < 6 \times 10^{-20}$ for all three GO branches) (Table 1; see also Supplementary Tables S2 and S5 and Supplementary Data XI).

The higher likelihood of LCI interactions to share other biological attributes is not surprising given inspection bias and potential circularity where functional annotation has been derived from an LCI interaction. That the CCSB-HI1 interaction pairs (which do not have such biases) yield statistically significant correlation supports their biological relevance. In all, 357 CCSB-HI1 interactions are supported by at least one additional characteristic and thus represent particularly appealing hypotheses of functional relatedness (Supplementary Fig. S4b). Lack of additional biological evidence is not an argument against any interaction (Supplementary Data XII). Importantly, complementary information from the interactome and other functional genomic data can be integrated to formulate biological models²⁵.

The CCSB-HI1 network has topological properties that are similar to other sampled interactome networks, such as an approximately power-law degree distribution, hierarchical organization and a tendency for highly connected (hub) proteins to interact with less highly connected proteins (Supplementary Data XIII and Supplementary Fig. S5a–d). Surprisingly, although the CCSB-HI1 network has a small characteristic path length, it does not exhibit high clustering, seemingly contradicting findings from the sparsely sampled networks of other organisms that protein interaction networks are ‘small world’^{6,7,26} (that is, have a short characteristic path length and a high clustering coefficient). Possible explanations for this apparent discrepancy are discussed in Supplementary Data XIII.

To gain an insight into the evolution of the interactome, we classified proteins in the CCSB-HI1 network as ‘eukaryotic’, ‘metazoan’, ‘mammalian’ or ‘human’, and asked whether proteins specific to different evolutionary classes tend to interact with one another. The CCSB-HI1 network appears to be enriched for interactions between proteins of the same evolutionary class but not for interactions between proteins from two different evolutionary classes (Supplementary Table S6). This suggests that the human interactome has evolved through the preferential addition of interactions between lineage-specific proteins. Further investigation may provide hypotheses for mechanisms underlying interactome evolution.

To detect densely connected subgraphs potentially representing biological modules, we applied the MCODE graph clustering algorithm²⁷ to the CCSB-HI1 and to the combined CCSB-HI1–LCI and CCSB-HI1–LC networks (Supplementary Fig. S6, Supplementary Table S7 and Supplementary Data XIV). We identified functionally enriched MCODE complexes using FuncAssociate²⁸. Out of 172 complexes (Supplementary Table S7) containing at least one CCSB-HI1 interaction, we identified 102 in which at least one GO term was significantly over-represented ($P \leq 0.05$), ten times the number expected by chance alone. The enriched functional terms we identified may also apply to unannotated proteins present in the complex (‘guilt by association’ predictions).

CCSB-HI1 represents a repository of novel biological hypotheses for genes implicated in human diseases. We compared all CCSB-HI1 proteins to the list of genes associated with human diseases in the Online Mendelian Inheritance in Man (OMIM) database and identified 424 interacting pairs for which at least one partner had been previously associated with a human disease (Supplementary Table S8). In a query of PubMed and Google Scholar, searching for gene symbols, 352 of the 424 interaction pairs appeared to be new based on the absence of any hit in either database. Along with 79 LC interactions (including LCI and non-binary LC interactions) among proteins in this space, the resulting network contains 484 interactions among 417 proteins (Fig. 3).

In one example, we found an interaction between RTN4, a neurite outgrowth inhibitor, and SPG21, the spastic paraplegia 21 protein. Mutations in SPG21 cause an autosomal recessive motor disorder called Mast syndrome. SPG21 protein localizes to intracellular endosomal/trans-Golgi transportation vesicles and is thought to function in protein transport and sorting. Although the function of its interacting partner, RNT4, remains elusive, RNT4 belongs to a family of proteins that localize to the endoplasmic reticulum and are markers of neuroendocrine differentiation²⁹. In addition, RNT4 shares two regulatory motifs with SPG21 that are conserved across mammalian genomes²² and may have a role in Mast syndrome. This and other examples (Supplementary Data XV) suggest that CCSB-HI1 can be used to connect biological processes in order to understand further network and disease relationships.

Although the CCSB-HI1 data set is far from comprehensive, and incomplete sampling limits conclusions regarding some network properties, our data provide useful hypotheses and can guide further studies of the expanded network. Currently, CCSB-HI1 is a static graph. Eventually the dynamics of the human interactome network will need to be considered to address where and when interactions take place and how they are regulated. The functional consequences of these physical interactions will also have to be studied to understand the logic of complex biological networks. Just as the first drafts of the human genome changed strategies for disease gene identification, the emerging human interactome will greatly further the understanding of human health and disease.

METHODS

The CCSB-HI1 data set was generated using a high-throughput version of the yeast two-hybrid system. First, all 8,100 cloned ORFs of the human ORFeome v1.1 were transferred from Entry clones to both AD and DB vectors by Gateway recombinational cloning. Resulting constructs were transformed in haploid yeast cells. After mating, diploids were tested on selective media for their ability to grow in an AD-Y-dependent manner. The identity of the interactors was determined after PCR amplification and sequencing of the AD and DB inserts from positive colonies. Resulting interactions were re-tested by the yeast two-hybrid system, individually, assessed for quality by co-affinity purification assays and analysed for correlation with other biological information. For a detailed description of the various methods, see Supplementary Data.

Received 21 July; accepted 8 September 2005.
Published online 28 September 2005.

1. Fromont-Racine, M., Rain, J. C. & Legrain, P. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nature Genet.* **16**, 277–282 (1997).
2. Walhout, A. J. et al. Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science* **287**, 116–122 (2000).
3. Uetz, P. et al. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623–627 (2000).
4. Ito, T. et al. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl Acad. Sci. USA* **98**, 4569–4574 (2001).
5. Reboul, J. et al. *C. elegans* ORFeome version 1.1: experimental verification of the genome annotation and resource for proteome-scale protein expression. *Nature Genet.* **34**, 35–41 (2003).
6. Giot, L. et al. A protein interaction map of *Drosophila melanogaster*. *Science* **302**, 1727–1736 (2003).
7. Li, S. et al. A map of the interactome network of the metazoan *C. elegans*. *Science* **303**, 540–543 (2004).

8. Jeong, H., Mason, S. P., Barabasi, A. L. & Oltvai, Z. N. Lethality and centrality in protein networks. *Nature* **411**, 41–42 (2001).
9. Han, J. D. *et al.* Evidence for dynamically organized modularity in the yeast protein–protein interaction network. *Nature* **430**, 88–93 (2004).
10. Vidal, M. A biological atlas of functional maps. *Cell* **104**, 333–339 (2001).
11. Xenarios, I. *et al.* DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res.* **30**, 303–305 (2002).
12. Zanzoni, A. *et al.* MINT: a Molecular INTERaction database. *FEBS Lett.* **513**, 135–140 (2002).
13. Bader, G. D., Betel, D. & Hogue, C. W. BIND: the Biomolecular Interaction Network Database. *Nucleic Acids Res.* **31**, 248–250 (2003).
14. Peri, S. *et al.* Development of human protein reference database as an initial platform for approaching systems biology in humans. *Genome Res.* **13**, 2363–2371 (2003).
15. Pagel, P. *et al.* The MIPS mammalian protein-protein interaction database. *Bioinformatics* **21**, 832–834 (2005).
16. Lehner, B. & Fraser, A. A first-draft human protein-interaction map. *Genome Biol.* **5**, R63 (2004).
17. Rual, J. F. *et al.* Human ORFeome version 1.1: a platform for reverse proteomics. *Genome Res.* **14**, 2128–2135 (2004).
18. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004).
19. Walhout, A. J. & Vidal, M. A genetic strategy to eliminate self-activator baits prior to high-throughput yeast two-hybrid screens. *Genome Res.* **9**, 1128–1134 (1999).
20. Gavin, A. C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141–147 (2002).
21. Formstecher, E. *et al.* Protein interaction mapping: A *Drosophila* case study. *Genome Res.* **15**, 376–384 (2005).
22. Xie, X. *et al.* Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* **434**, 338–345 (2005).
23. Eppig, J. T. *et al.* The Mouse Genome Database (MGD): from genes to mice—a community resource for mouse biology. *Nucleic Acids Res.* **33**, D471–D475 (2005).
24. Camon, E. *et al.* The Gene Ontology Annotation (GOA) Database: sharing knowledge in Uniprot with Gene Ontology. *Nucleic Acids Res.* **32**, D262–D266 (2004).
25. Gunsalus, K. C. *et al.* Predictive models of molecular machines involved in *Caenorhabditis elegans* early embryogenesis. *Nature* **436**, 861–865 (2005).
26. Wagner, A. The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes. *Mol. Biol. Evol.* **18**, 1283–1292 (2001).
27. Bader, G. D. & Hogue, C. W. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* **4**, 2 (2003).
28. Berriz, G. F., King, O. D., Bryant, B., Sander, C. & Roth, F. P. Characterizing gene sets with FuncAssociate. *Bioinformatics* **19**, 2502–2504 (2003).
29. Huang, X. *et al.* Overexpression of human reticulon 3 (hRTN3) in astrocytoma. *Clin. Neuropathol.* **23**, 1–7 (2004).
30. Johnson, J. M. *et al.* Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* **302**, 2141–2144 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This paper is dedicated to the memory of Stan Korsmeyer. We thank members of the Vidal laboratory and the participants of the ORFeome Meeting for discussions; the sequencing staff at Agencourt Biosciences for technical assistance; E. Smith for his help with the figures; C. McCowan, A. Bird, T. Clingingsmith and C. You for administrative assistance; and E. Benz, S. Korsmeyer, D. Livingston, P. McCue, J. Song, B. Rollins and the DFCI Strategic Planning Initiative for support. Our human interactome project is supported by the DFCI High-Tech Fund (S. Korsmeyer), an Ellison Foundation grant awarded to M.V., an NIH/NCI grant awarded to S. Korsmeyer, S. Orkin, G. Gilliland and M.V., an 'interactome mapping' grant from NIH/NHGRI and NIH/NIGMS awarded to F.P.R. and M.V., and a W.M. Keck Foundation grant awarded to E. Benz, J. Marto, F.P.R. and M.V. Other support includes Taplin Funds for Discovery (F.P.R., F.D.G. and G.F.B.), a 2003 NSF Fellowship (D.S.G.) and funding from the Fonds National de la Recherche Scientifique, Belgium (M.D.).

Author Contributions Experiments and data analyses were coordinated by J.F.R., T.H. and K.V. High-throughput ORF cloning and yeast two-hybrid screens were performed by J.F.R., T.H.K., A.D., N.L., N.A.G., J.R. and J.L. J.F.R. developed the high-throughput yeast two-hybrid strategy. Computational analyses were performed by T.H., K.V., G.F.B., F.D.G., N.K., P.L., D.S.G., L.V.Z., S.L.W. and G.F. Co-affinity purification experiments were performed by M.D., C.S., J.F.R., S.M., M.B., S.L. and J.S.A. C.F., E.L., S.C. and C.B. provided laboratory support. R.S.S., J.V., H.Y.Z., A.S. and M.E.C. helped with the overall interpretation of the data. DNA sequencing was performed by S.B., R.S. and L.D.S. The manuscript was written by J.F.R., K.V., M.E.C., D.E.H., F.P.R. and M.V. The project was conceived by M.V. and co-directed by D.E.H., F.P.R. and M.V.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.V. (marc_vidal@dfci.harvard.edu), F.P.R. (fritz_roth@hms.harvard.edu) or D.E.H. (david_hill@dfci.harvard.edu).

Photosystem II core phosphorylation and photosynthetic acclimation require two different protein kinases

Vera Bonardi^{1,2*}, Paolo Pesaresi^{2,3*}, Thomas Becker¹, Enrico Schleiff¹, Raik Wagner⁴, Thomas Pfannschmidt⁴, Peter Jahns⁵ & Dario Leister^{1,2}

Illumination changes elicit modifications of thylakoid proteins and reorganization of the photosynthetic machinery. This involves, in the short term, phosphorylation of photosystem II (PSII) and light-harvesting (LHCII) proteins. PSII phosphorylation is thought to be relevant for PSII turnover^{1,2}, whereas LHCII phosphorylation is associated with the relocation of LHCII and the redistribution of excitation energy (state transitions) between photosystems^{3,4}. In the long term, imbalances in energy distribution between photosystems are counteracted by adjusting photosystem stoichiometry^{5,6}. In the green alga *Chlamydomonas* and the plant *Arabidopsis*, state transitions require the orthologous protein kinases STT7 and STN7, respectively^{7,8}. Here we show that in *Arabidopsis* a second protein kinase, STN8, is required for the quantitative phosphorylation of PSII core proteins. However, PSII activity under high-intensity light is affected only slightly in *stn8* mutants, and D1 turnover is indistinguishable from the wild type, implying that reversible protein phosphorylation is not essential for PSII repair. Acclimation to changes in light quality is defective in *stn7* but not in *stn8* mutants, indicating that short-term and long-term photosynthetic adaptations are coupled. Therefore the phosphorylation of LHCII, or of an unknown substrate of STN7, is also crucial for the control of photosynthetic gene expression.

STT7 and STN7 are orthologous protein kinases required for LHCII phosphorylation and for state transitions in *Chlamydomonas* and *Arabidopsis*, respectively^{7,8}. In *Arabidopsis*, another STT7/STN7-like protein (STN8) exists that is not required for state transitions⁸. STN8 is located in the chloroplast, as shown by *in vivo* subcellular localization of its amino-terminal region fused to the dsRED protein and by the import of, and transit peptide removal from, STN8 translated *in vitro* (Fig. 1a, b). Chloroplast subfractionation after import revealed that the protein is associated, like STT7 and STN7, with thylakoids (Fig. 1c) (refs 7, 8).

Insertion mutants for *STN8* and *STN7* were obtained from the Salk collection⁹, and for each gene two independent mutant alleles lacking the respective transcript were identified (Supplementary Fig. S1). The *stn7 stn8* double mutant was generated by crossing *stn7* and *stn8* single knockouts and screening the resulting F₂ generation for homozygous double mutants. All mutants were indistinguishable from the wild type with regard to the timing of seed germination and growth rate in the greenhouse (Supplementary Fig. S1). In *stn7* and *stn7 stn8* mutants, a slight decrease in the levels of neoxanthin, lutein and total chlorophyll was found (Supplementary

Table S1). These subtle changes can be attributed to a minor decrease in LHCII content, not detectable by polyacrylamide-gel electrophoresis (PAGE) analysis (Supplementary Fig. S2).

Photosynthetic electron flow, measured on the basis of chlorophyll fluorescence, was not altered in the mutants (Supplementary Table S2). State transitions were suppressed in *stn7* and *stn7 stn8* plants (Supplementary Table S2) but were not affected in the *stn8* mutant, confirming previous results⁸. Reversible LHCII phosphorylation, which is associated with state transitions⁷, was studied *in vivo* in dark-adapted plants incubated with [³³P]orthophosphate and then exposed to different light conditions (Fig. 2a). Under low light, wild-type and *stn8* plants showed a marked increase in LHCII phosphorylation, whereas subsequent exposure to high light decreased the amount of phospho-LHCII. In *stn7* and *stn7 stn8* plants, reversible phosphorylation of LHCII was not detectable, again in agreement with previous results⁸. Determination of LHCII phosphorylation with a phosphothreonine-specific antibody (Fig. 2b), or an *in vitro* assay in conditions under which the LHCII

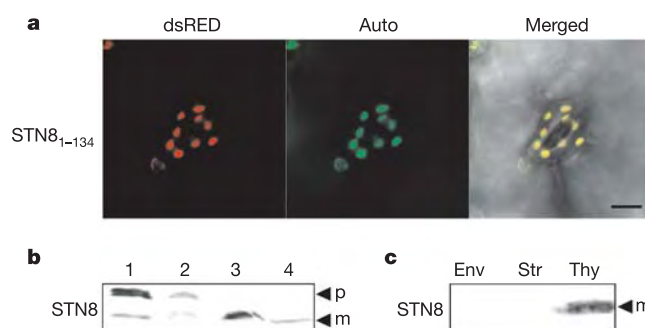


Figure 1 | Subcellular localization of STN8. **a**, The STN8₍₁₋₁₃₄₎-dsRED fusion was stably introduced into *stn8* plants. Guard cells were analysed by confocal laser scanning microscopy. Left, dsRED fluorescence identifying the fusion protein; middle, chloroplasts revealed by chlorophyll autofluorescence (shown in false colour); right, merged images. Scale bar, 50 μ m. **b**, ³⁵S-labelled protein, translated *in vitro* (lane 1, 10% translation product) was incubated with isolated chloroplasts for 20 min at 4 °C (lane 2) or 25 °C (lanes 3 and 4), and chloroplasts were recovered by centrifugation through 40% Percoll. Chloroplasts were incubated with thermolysin (lane 4) and subjected to SDS-PAGE; proteins were detected by autoradiography. p, precursor; m, mature protein. **c**, Chloroplasts were fractionated after protein import. Env, envelope; Str, stroma; Thy, thylakoids.

¹Botanisches Institut, Department Biologie I, Ludwig-Maximilians-Universität, Menzinger Strasse 67, 80638 München, Germany. ²Abteilung für Pflanzenzüchtung und Genetik, Max-Planck-Institut für Züchtungsforschung, Carl-von-Linné-Weg 10, 50829 Köln, Germany. ³Fondazione Parco Tecnologico Padano, Via Albert Einstein Cascina Codazza, 26900 Lodi, Italy. ⁴Lehrstuhl für Pflanzenphysiologie, Friedrich-Schiller-Universität Jena, Dornburger Strasse 159, 07743 Jena, Germany. ⁵Institut für Biochemie der Pflanzen, Heinrich-Heine-Universität Düsseldorf, Universitätsstrasse 1, 40225 Düsseldorf, Germany.

*These authors contributed equally to this work.

kinase should be maximally active (Fig. 2c), also showed that a loss of STN7 function, but not of STN8 function, suppresses the phosphorylation of LHCII. These results, together with the spectroscopic data, indicate that STN8 and STN7 do not act in series in phosphorylating LHCII.

Thylakoid protein accumulation was similar in mutant and wild-type plants (Supplementary Fig. S2). When phosphorylation of PSII core proteins was monitored with the phosphothreonine-specific antibody, *stn7* plants behaved like the wild type, showing high levels of phosphorylated PSII core proteins under all light regimes whereas LHCII phosphorylation was greatly decreased (Fig. 2d, e). In contrast, *stn8* plants showed a marked decrease in the total amount of PSII core phosphoproteins—particularly under high light—whereas LHCII phosphorylation was as in the wild type. This again argues that the main substrates of STN7 and STN8 are different: LHCII phosphorylation is mostly dependent on STN7, whereas phosphorylation of PSII core proteins depends almost exclusively on STN8. Strikingly, only in the *stn7 stn8* mutant were the phosphorylated forms of LHCII and the PSII core proteins completely absent under all light regimes tested, indicating that the two kinases must show some degree of overlap in their substrate specificities. However, the clear distinction between the phosphorylation phenotypes of the two mutants implies that STN7 and STN8 act in parallel and could be directly responsible for phosphorylating the LHCII and PSII core proteins, respectively. If, however, the proposal that LHCII is

phosphorylated by thylakoid-associated kinase (TAK) proteins¹⁰ is correct, then STN7 might act upstream of TAKs.

Reversible phosphorylation of the D1 protein is thought to have a key function in the regulation of its turnover during the photo-inhibition of PSII (ref. 11). Exposure to light induces phosphorylation of, and causes damage to, PSII reaction centres, and the phosphorylated form of damaged D1 is resistant to proteolysis¹². However, relocation of damaged PSII centres from grana to stroma lamellae permits the dephosphorylation and proteolysis of D1, and co-translational incorporation of newly synthesized D1 (ref. 1). Consequently, lack of the D1 protease impairs the PSII repair cycle¹³. To test whether suppression of PSII core phosphorylation also impairs PSII repair by changing the rate of D1 turnover, we investigated the inactivation of PSII under high light, the recovery of PSII activity and the degradation of D1. Illumination of leaves at high light intensity led to a slightly stronger inactivation of PSII in *stn8* and *stn7 stn8* plants than in *stn7* and wild-type plants; the subsequent recovery of PSII activity under low light was also somewhat slower in *stn8* and *stn7 stn8* than in wild-type and *stn7* plants (Fig. 3a). However, the increased photosensitivity of PSII in *stn8* and *stn7 stn8* plants was not reflected in changes in the rate of light-induced D1 degradation in the presence of lincomycin, an inhibitor of plastid protein synthesis (Supplementary Fig. S3). Moreover, also during photoinhibition and subsequent recovery, no PSII core phosphorylation was detected in *stn7 stn8* plants (Supplementary Fig. S4), excluding the action of another PSII core kinase under these conditions. Pulse-chase experiments under high light intensities

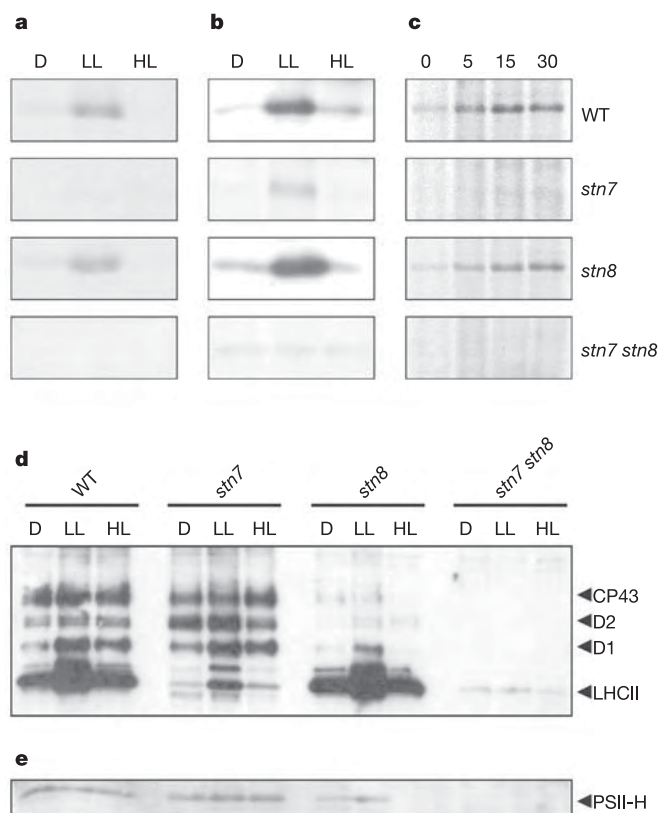


Figure 2 | Phosphorylation of thylakoid proteins. **a**, *In vivo* LHCII phosphorylation. Leaves were incubated with [³³P]orthophosphate, and thylakoid proteins from plants kept in the dark (D), subsequently exposed to low light (LL), and then to high light (HL), were fractionated by SDS-PAGE. WT, wild type. **b**, LHCII phosphorylation detected by immunoblot analysis with a phosphothreonine-specific antibody. Leaves were treated and proteins fractionated as in **a**. **c**, *In vitro* LHCII phosphorylation. Thylakoids from dark-adapted leaves were incubated with [³³P]ATP under reducing conditions for 0, 5, 15 and 30 min in the dark. Proteins were fractionated as in **a**. **d**, **e**, Phosphorylation of LHCII and PSII core proteins detected by immunoblot analysis. Leaves were treated and thylakoid proteins fractionated as in **b**.

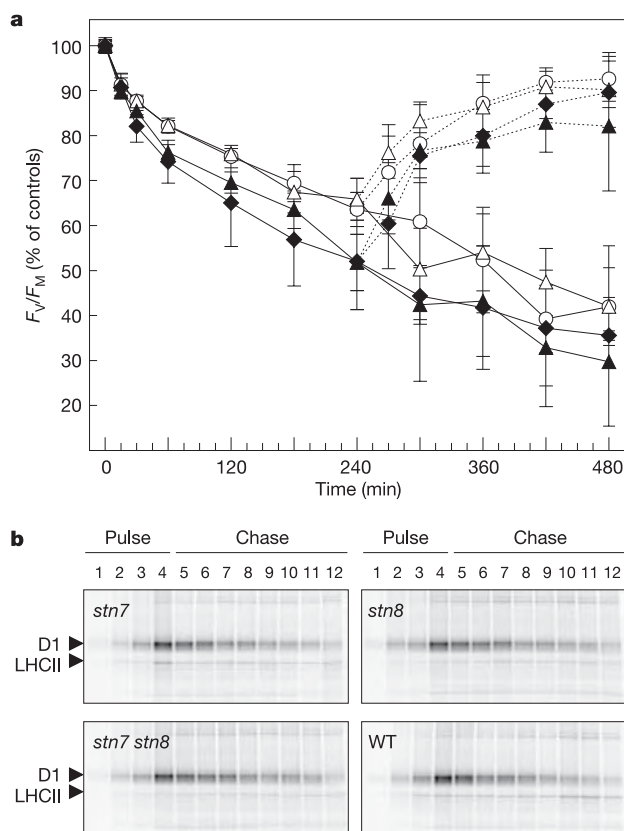


Figure 3 | PSII activity and D1 turnover under high-intensity light. **a**, Time course of PSII inactivation (solid lines) induced by high light ($2,000 \mu\text{mol photons m}^{-2} \text{s}^{-1}$) and recovery under low light ($20 \mu\text{mol photons m}^{-2} \text{s}^{-1}$; dashed lines). Error bars indicate s.d. Open triangles, *stn7*; filled triangles, *stn8*; diamonds, *stn7 stn8*; circles, wild type. **b**, Autoradiogram of thylakoid membrane proteins resolved by SDS-PAGE after pulse-labelling with [³⁵S]methionine for 0, 15, 30 and 60 min (lanes 1–4) and subsequent chase in unlabelled medium for 60, 120, 180, 240, 300, 360, 420 and 480 min (lanes 5–12), both under high light ($2,000 \mu\text{mol photons m}^{-2} \text{s}^{-1}$). WT, wild type.

revealed that rates of D1 synthesis and degradation were similar in all genotypes (Fig. 3b; Supplementary Fig. S5). In addition, wild-type, *stn8* and *stn7 stn8* plants grown under high light intensities behaved very similarly in growth rate and leaf pigment composition (data not shown). Taken together, the data indicate that STN8-mediated phosphorylation of D1 is not crucial for D1 turnover and PSII repair.

Changes in light conditions are thought to result, in the long term, in the adjustment of photosystem stoichiometry, which requires a signalling network that coordinates photosynthetic gene expression in plastids and nucleus^{5,6,14}. Recent studies have proposed a functional relationship between LHCII phosphorylation (and state transitions) on the one hand, and the long-term response to altered light conditions on the other^{15,16}. The slower growth of *stn7* mutants under fluctuating light has been attributed to impaired state-transition-based adaptation⁸. Here we tested whether defects in the function of STN7 or STN8 affect the changes in antenna structure and photosystem stoichiometry that are associated with the long-term response to changes in the quality of incident light^{17,18}. Plants were grown and acclimated to light sources that favoured either PSI or PSII, and the long-term response was followed by monitoring the chlorophyll fluorescence parameter F_S/F_M and the ratio of chlorophyll *a* to chlorophyll *b* (Chl *a/b*). In the wild type, F_S/F_M values were relatively high after acclimation to PSI light, and low values were measured

after acclimation to PSII light, whereas the Chl *a/b* ratio behaved in the opposite sense, being high after acclimation to PSII light and low in PSI light^{17,18} (Fig. 4a, b). The *stn8* mutant behaved like the wild type (Fig. 4a, b), indicating that its long-term response was normal. In the *stn7* and *stn7 stn8* mutants, however, F_S/F_M and Chl *a/b* values were typical of plants acclimated to PSI light under all light regimes tested, indicating that the mutants have lost their capacity for the long-term response and implying that STN7 has a function in coordinating the long-term and short-term responses to changes in light conditions.

Lack of STN7 might impair the long-term response by interfering with a signalling pathway that links changes in photosynthetic efficiency during the adjustment of photosystem stoichiometry to the level of gene expression in plastids¹⁹ and the nucleus^{17,18}. To test this we analysed the transcription of photosynthetic genes. For nuclear genes encoding chloroplast proteins two modes of transcriptional regulation are known: a master switch that acts on most genes²⁰ and an additional mechanism specific for photosynthetic genes²¹. Lack of STN7 in greenhouse-grown *stn7* and *stn7 stn8* plants results in the differential expression of only relatively few photosynthetic genes, in contrast to *stn8* plants in which a large set of photosynthetic genes is markedly downregulated (Fig. 4c; Supplementary Table S3); exemplary expression profiles of two nuclear and two plastid genes are provided in Fig. 4d. The impairment of the transcriptional regulation of certain *stn8*-responsive genes in the absence of STN7, together with the results of the long-term response experiments, therefore argues in favour of a function for STN7 in the regulation of nuclear and plastid gene expression. We can only speculate how STN7 triggers changes in photosynthetic gene expression but, in principle, three hypotheses are available: first, the phosphorylation state of LHCII directly provides information for signalling; second, an unknown protein is phosphorylated by STN7 and participates in signalling; and third, state transitions and the associated conformational changes of thylakoids^{3,22} stimulate signalling.

In green algae, the pool of mobile LHCII is large²³, and state transitions are important for cyclic electron flow³. Because the mobile LHCII pool in vascular plants is relatively small²⁴, it is tempting to speculate that triggering of the long-term response, rather than the short-term response in terms of state transitions, represents the major function of STN7 in flowering plants. In this respect, the slower growth of *stn7* plants under fluctuating light⁸ might be due to disturbance of transcriptional regulation rather than being a physiological consequence of defects in state transitions. Future analyses must clarify how the short-term and long-term responses are coupled, and whether STN7 and STN8 are necessary and sufficient for the phosphorylation of thylakoid proteins.

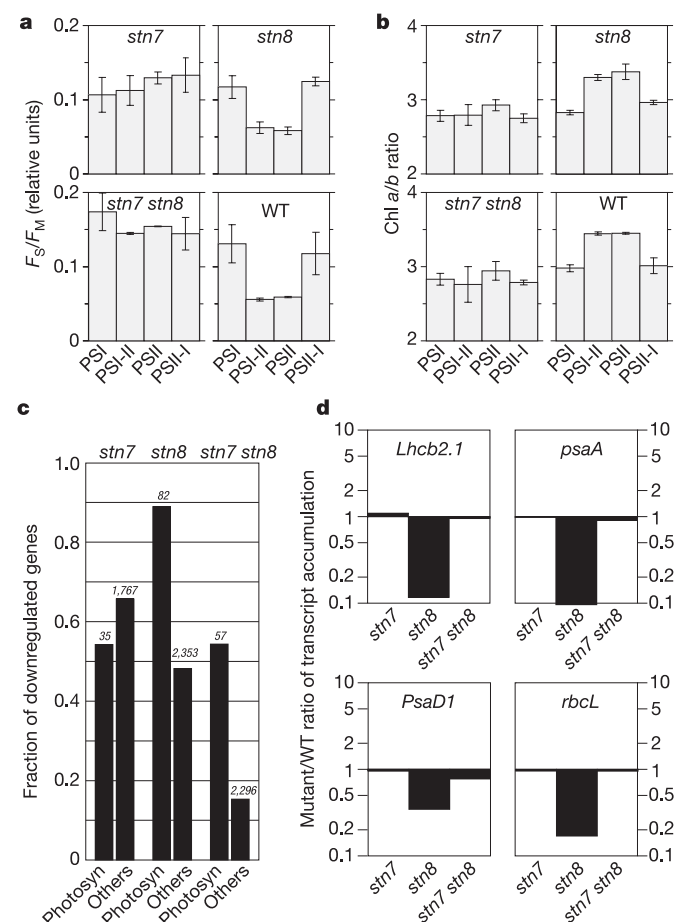


Figure 4 | Photosynthetic acclimation and mRNA expression. **a, b**, Plants were acclimated to PSI or PSII light and F_S/F_M values (**a**) and Chl *a/b* ratios (**b**) were determined. Error bars indicate s.d. WT, wild type. **c**, Nuclear transcript accumulation in greenhouse-grown plants. Numbers above bars indicate the total numbers of genes that are significantly differentially expressed in mutants with respect to wild-type plants. Bar lengths indicate the fraction of differentially regulated genes that were downregulated. Photosyn, genes for photosynthesis; others, other genes (for non-photosynthetic proteins). **d**, Mutant versus wild-type ratios of transcript levels (logarithmic scale) of selected nuclear and plastid photosynthetic genes.

METHODS

Plant lines and propagation. Mutant lines from the Salk collection⁹ were identified by searching the SiGNAL database (<http://signal.salk.edu/about.html>). Methods for plant propagation have been described elsewhere^{25–27}.

Chlorophyll fluorescence and pigment analysis. Photosynthetic electron transport, state transitions and leaf pigment composition were measured as described^{25–27} (see Supplementary Information for details).

Analysis of LHCII phosphorylation *in vivo* and *in vitro*. To determine the degree of LHCII phosphorylation *in vivo*, dark-adapted leaves from 4-week-old plants were incubated in the presence of [³³P]orthophosphate for 1 h and subsequently exposed for 2 h to levels of illumination favouring phosphorylation (low light, 80 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$) or dephosphorylation (high light, 800 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$). Identical amounts of thylakoid proteins (equivalent to 200 mg of fresh leaf) were prepared in the presence of 10 mM NaF²⁵ and fractionated by SDS-PAGE (14% polyacrylamide); labelled proteins were detected by phosphorimaging (Typhoon; Amersham Biosciences). For the assay *in vitro*, thylakoids isolated from dark-adapted leaves (30 mg of fresh leaf) were incubated with [γ -³³P]ATP under reducing conditions in the dark²⁵. Separation and detection of thylakoid proteins were performed as described for the *in vivo* assay.

Immunoblot analysis. Leaves from 4-week-old plants were harvested after overnight dark adaptation or light exposure as described above. Thylakoids

were prepared in the presence of 10 mM NaF, fractionated on an SDS–polyacrylamide-gradient gel (8–25% polyacrylamide) and transferred to poly(vinylidene difluoride) membranes²⁵. Filters were then probed with antibodies specific for phosphothreonine (Biolabs) and signals were detected by enhanced chemiluminescence (Amersham Biosciences).

Pulse–chase measurement of D1 turnover. The pulse–chase procedure for the analysis of D1 turnover in pea²⁸ was modified for *Arabidopsis*. For radioactive labelling of thylakoid proteins, leaf discs of 3-week-old *Arabidopsis* plants were pressed extremely gently against coarse sand paper and then vacuum-infiltrated in a syringe containing 1 mCi of L-[³⁵S]methionine in 10 ml of 1 mM KH₂PO₄ pH 6.3, 0.1% Tween 20. Directly after infiltration, three leaf discs were frozen in nitrogen (t_0). Remaining leaves were transferred to high light (2,000 μ mol photons $m^{-2} s^{-1}$) and for each time point ($t_{pulse} = 15, 30$ and 60 min) three leaf discs were collected. Immediately after the pulse period, the remaining leaf discs were washed, incubated with 10 mM unlabelled L-methionine in the same buffer as before and further exposed to high light for up to 8 h ($t_{chase} = 60, 120, 180, 240, 300, 360, 420$ and 480 min). The three leaf discs for each time point were combined and thylakoid proteins were prepared, separated and detected as described for the LHCII phosphorylation analysis.

In vitro import and intracellular localization of dsRED fusions in chloroplasts. For *in vitro* import assays, ³⁵S-labelled proteins were synthesized and used for import experiments and were detected after subfractionation of chloroplasts^{29,30}. For intracellular localization of the dsRED fusion, a complementary DNA fragment coding for the first 134 amino-acid residues of STN8 was fused 5' to *dsRED* and inserted into the vector pLEELA (Invitrogen), placing it under the transcriptional control of the cauliflower mosaic virus 35S promoter. Seeds were collected from transformed *stn8* plants²⁶ and independent transgenic plants were selected. Confocal images were collected by laser scanning microscopy (TCS SP2; Leica). Fluorescence was excited with a 461 nm HeNe laser and images were collected in the ranges 565–620 nm (dsRED fluorescence) and 670–750 nm (chlorophyll autofluorescence).

Nucleic acid analysis. *Arabidopsis* DNA was isolated²⁶ and T-DNA insertion junction sites were recovered by polymerase chain reaction (PCR) with the use of combinations of insertion-specific and gene-specific primers, and then sequenced. To determine levels of gene expression, extraction of total leaf RNA, first-strand cDNA synthesis and reverse-transcriptase-mediated PCR (RT–PCR) were performed²⁶, using primers specific for *STN7* or *STN8*, as well as *ACTIN1*-specific oligonucleotides as a control.

mRNA expression profiling. Greenhouse-grown mutant and wild-type plants were analysed. The generation and use of a 3292-gene sequence tag (GST) nylon array enriched for nuclear genes for chloroplast proteins, data analysis and statistical evaluation have all been described previously^{20,21,26} (see Supplementary Information for details).

Measurement of acclimation to changes in light quality. Light conditions favouring either PSI or PSII (PSI or PSII light, respectively), as well as growth and acclimation conditions for *Arabidopsis*, have been described previously¹⁸. Plants were initially grown for 10 days under white light followed by a 6-day acclimation period. Seedlings were acclimated either to PSI or PSII light for 6 days or to PSI light for 2 days followed by 4 days under PSII light or vice versa. Measurement of Chl *a/b* ratios and determination of the chlorophyll fluorescence parameters F_S (steady-state fluorescence) and F_M (maximal fluorescence), and of F_S/F_M ratios, were performed as described in Supplementary Information. All values under the four conditions were calculated as the means of 50 individuals in at least three independent experiments, and the significance of differences between samples was tested with Student's *t*-test.

Received 3 May; accepted 11 July 2005.

1. Rintamäki, E., Kettunen, R. & Aro, E.-M. Differential D1 dephosphorylation in functional and photodamaged photosystem II centers. Dephosphorylation is a prerequisite for degradation of damaged D1. *J. Biol. Chem.* **271**, 14870–14875 (1996).
2. Rintamäki, E. *et al.* Phosphorylation of light-harvesting complex II and photosystem II core proteins shows different irradiance-dependent regulation *in vivo*—application of phosphothreonine antibodies to analysis of thylakoid phosphoproteins. *J. Biol. Chem.* **272**, 30476–30482 (1997).
3. Wollman, F. A. State transitions reveal the dynamics and flexibility of the photosynthetic apparatus. *EMBO J.* **20**, 3623–3630 (2001).
4. Allen, J. F. & Forsberg, J. Molecular recognition in thylakoid structure and function. *Trends Plant Sci.* **6**, 317–326 (2001).
5. Fujita, Y. A study on the dynamic features of photosystem stoichiometry: Accomplishments and problems for future studies. *Photosynth. Res.* **53**, 83–93 (1997).
6. Melis, A. Dynamics of photosynthetic membrane composition and function. *Biochim. Biophys. Acta* **1058**, 87–106 (1991).
7. Depege, N., Bellafiore, S. & Rochaix, J. D. Role of chloroplast protein kinase Stt7

- in LHCII phosphorylation and state transition in *Chlamydomonas*. *Science* **299**, 1572–1575 (2003).
8. Bellafiore, S., Barneche, F., Peltier, G. & Rochaix, J. D. State transitions and light adaptation require chloroplast thylakoid protein kinase STN7. *Nature* **433**, 892–895 (2005).
9. Alonso, J. M. *et al.* Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* **301**, 653–657 (2003).
10. Snyders, S. & Kohorn, B. D. TAKS, thylakoid membrane protein kinases associated with energy transduction. *J. Biol. Chem.* **274**, 9137–9140 (1999).
11. Baena-González, E., Barbato, R. & Aro, E.-M. Role of phosphorylation in the repair cycle and oligomeric structure of photosystem II. *Planta* **208**, 196–204 (1999).
12. Koivuniemi, A., Aro, E.-M. & Andersson, B. Degradation of the D1- and D2-proteins of photosystem II in higher plants is regulated by reversible phosphorylation. *Biochemistry* **34**, 16022–16029 (1995).
13. Bailey, S. *et al.* A critical role for the Var2 FtsH homologue of *Arabidopsis thaliana* in the photosystem II repair cycle *in vivo*. *J. Biol. Chem.* **277**, 2006–2011 (2002).
14. Pfannschmidt, T. Chloroplast redox signals: how photosynthesis controls its own genes. *Trends Plant Sci.* **8**, 33–41 (2003).
15. Pursiainen, S., Mulo, P., Rintamäki, E. & Aro, E.-M. Coregulation of light-harvesting complex II phosphorylation and *lhcb* mRNA accumulation in winter rye. *Plant J.* **26**, 317–327 (2001).
16. Allen, J. F. & Pfannschmidt, T. Balancing the two photosystems: photosynthetic electron transfer governs transcription of reaction centre genes in chloroplasts. *Phil. Trans. R. Soc. Lond. B* **355**, 1351–1359 (2000).
17. Pfannschmidt, T., Schütze, K., Brost, M. & Oelmüller, R. A novel mechanism of nuclear photosynthesis gene regulation by redox signals from the chloroplast during photosystem stoichiometry adjustment. *J. Biol. Chem.* **276**, 36125–36130 (2001).
18. Fey, V. *et al.* Retrograde plastid redox signals in the expression of nuclear genes for chloroplast proteins of *Arabidopsis thaliana*. *J. Biol. Chem.* **280**, 5318–5328 (2005).
19. Pfannschmidt, T., Nilsson, A. & Allen, J. F. Photosynthetic control of chloroplast gene expression. *Nature* **397**, 625–628 (1999).
20. Richly, E. *et al.* Covariations in the nuclear chloroplast transcriptome reveal a regulatory master-switch. *EMBO Rep.* **4**, 491–498 (2003).
21. Bieh, A., Richly, E., Noutsos, C., Salamini, F. & Leister, D. Analysis of 101 nuclear transcriptomes reveals 23 distinct regulons and their relationship to metabolism, chromosomal gene distribution and co-ordination of nuclear and plastid gene expression. *Gene* **344**, 33–41 (2005).
22. Vallon, O. *et al.* Lateral redistribution of cytochrome *b₆/f* complexes along thylakoid membranes upon state transitions. *Proc. Natl Acad. Sci. USA* **88**, 8262–8266 (1991).
23. Delosme, R., Beal, D. & Joliet, P. Photoacoustic detection of flash-induced charge separation in photosynthetic systems—spectral dependence of the quantum yield. *Biochim. Biophys. Acta* **1185**, 56–64 (1994).
24. Allen, J. F. Protein phosphorylation in regulation of photosynthesis. *Biochim. Biophys. Acta* **1098**, 275–335 (1992).
25. Pesaresi, P. *et al.* A stable LHCII-PSI aggregate and suppression of photosynthetic state transitions in the *psae1-1* mutant of *Arabidopsis thaliana*. *Planta* **215**, 940–948 (2002).
26. Ihnatowicz, A. *et al.* Mutants for photosystem I subunit D of *Arabidopsis thaliana*: effects on photosynthesis, photosystem I stability and expression of nuclear genes for chloroplast functions. *Plant J.* **37**, 839–852 (2004).
27. Grasses, T. *et al.* The role of Δ pH-dependent dissipation of excitation energy in protecting photosystem II against light-induced damage in *Arabidopsis thaliana*. *Plant Physiol.* **140**, 41–49 (2002).
28. Aro, E. M., McCaffery, S. & Andersson, J. M. Photoinhibition and D1 protein degradation in peas acclimated to different growth irradiances. *Plant Physiol.* **103**, 835–843 (1993).
29. Qbadou, S., Tien, R., Soll, J. & Schleiff, E. Membrane insertion of the chloroplast outer envelope protein, Toc34: constrains for insertion and topology. *J. Cell Sci.* **116**, 837–846 (2003).
30. Schleiff, E., Soll, J., Küchler, M., Kühlbrandt, W. & Harrer, R. Characterization of the translocon of the outer envelope of chloroplasts. *J. Cell Biol.* **160**, 541–551 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Bieh, C. Noutsos, A. Dietzmann, B. Eilts and M. Brost for excellent technical assistance; F. Salamini, M. Koornneef and P. Hardy for critical comments on the manuscript; and the Salk Institute for making T-DNA insertion lines publicly available. This work was supported by the European Community (Human Potential Programme (PSICO)) and the Deutsche Forschungsgemeinschaft.

Author Information Complete data sets are deposited at GEO (<http://www.ncbi.nlm.nih.gov/geo/>) under accession numbers GSE2620–GSE2622. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.L. (leister@lrz.uni-muenchen.de).

Crystal structure of a junction between B-DNA and Z-DNA reveals two extruded bases

Sung Chul Ha¹, Ky Lowenhaupt³, Alexander Rich³, Yang-Gyun Kim⁴ & Kyeong Kyu Kim^{1,2}

Left-handed Z-DNA is a higher-energy form of the double helix, stabilized by negative supercoiling generated by transcription or unwrapping nucleosomes¹. Regions near the transcription start site frequently contain sequence motifs favourable for forming Z-DNA², and formation of Z-DNA near the promoter region stimulates transcription^{3,4}. Z-DNA is also stabilized by specific protein binding; several proteins have been identified with low nanomolar binding constants^{5–9}. Z-DNA occurs in a dynamic state, forming as a result of physiological processes then relaxing to the right-handed B-DNA¹. Each time a DNA segment turns into Z-DNA, two B–Z junctions form. These have been examined extensively^{10–12}, but their structure was unknown. Here we describe the structure of a B–Z junction as revealed by X-ray crystallography at 2.6 Å resolution. A 15-base-pair segment of DNA is stabilized at one end in the Z conformation by Z-DNA binding proteins, while the other end remains B-DNA. Continuous stacking of bases between B-DNA and Z-DNA segments is found, with the breaking of one base pair at the junction and extrusion of the bases on each side (Fig. 1). These extruded bases may be sites for DNA modification.

A 15-base-pair DNA duplex was designed with two overhanging nucleotides at each end. We have taken advantage of the tight binding of the Z-DNA binding domain hZα_{ADAR1} from the human editing enzyme, double-stranded (ds)RNA adenosine deaminase (ADAR1)^{5,13}, to stabilize the Z conformation in one-half of the DNA duplex while the remainder of the DNA duplex retains the B conformation. One B–Z junction is thus formed in the middle of the DNA duplex, connecting the Z- and B-DNA. The hZα_{ADAR1} domain (amino acids 140–202) was co-crystallized with duplex DNA (5′-GTCGCGCGCCATAAACC-3′ and 5′-ACGTTTATGGCGC GCG-3′). The crystal has space group *P*₆₁, and the structure of the complex was solved at 2.6 Å resolution (*R* = 23.8%, *R*_{free} = 28.5%; Supplementary Table 1 and Supplementary Methods).

Two hZα_{ADAR1} peptides bind to each strand of the DNA duplex. Thus, four hZα_{ADAR1} peptides (ZαI–ZαIV) form a complex with one DNA duplex in the crystallographic asymmetric unit (Figs 1 and 2a, and Supplementary Fig. S1). The structures of the four hZα_{ADAR1} domains bound to DNA (ZαI–ZαIV) are nearly identical to that previously described for hZα_{ADAR1} (ref. 5) (Supplementary Fig. S2, Supplementary Table 2 and Supplementary Discussion). The 1:4 stoichiometry between the DNA duplex and hZα_{ADAR1} was confirmed by circular dichroism (CD) measurements (Supplementary Fig. S3). In the DNA duplex, eight base pairs are in the left-handed Z conformation, and six base pairs retain the right-handed B conformation (Figs 1 and 2a). Between the B and Z conformations, an A–T base pair is disrupted and one base from each strand (T0′ and A0) is extruded from the double helix (Figs 1 and 2b; note that base nomenclature is defined in Fig. 2 and its legend). The sugar

phosphate backbone of these bases links the B- and Z-DNA segments (Fig. 3); a sharp turn accommodates a reversal in the backbone direction. At the B–Z junction the DNA is bent by 11.05°, and the helical axes are displaced from each other by 5.2 Å, when calculated on the base plane of C–1 and G1′ (Fig. 3b).

The structure of the eight-base-pair region in the Z conformation is essentially the same as the previously described Z-DNA structure¹⁴ (Supplementary Table 3), with a zig-zag sugar phosphate backbone and an alternating *anti-syn* base conformation. The other end of the DNA duplex adopts a standard B conformation¹⁵. No interactions are found between hZα_{ADAR1} and the extruded bases (Fig. 2a), and the glycosyl bonds of A0 and T0′ are both in the *anti* conformation. The base of T0′ is oriented almost parallel to the helix axis, whereas the A0 base is fully extended out from the helix (Figs 1 and 3, and Supplementary Fig. S1). A segment of an electron density omit map for the extruded thymine T0′ and adenine A0 are shown in Fig. 2b and in Supplementary Fig. S4. The extruded adenine A0 has a

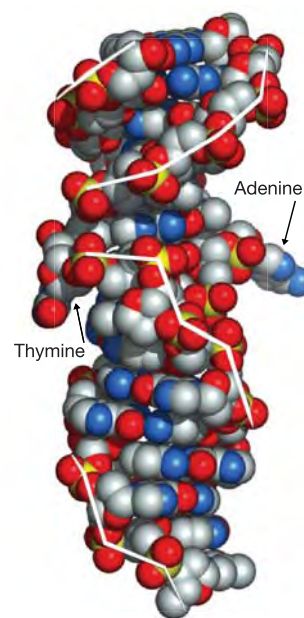


Figure 1 | A van der Waals view of the 15-base-pair DNA structure containing a junction between left-handed Z-DNA and right-handed B-DNA. Two bases have been extruded from base stacking at the junction. The white line goes from phosphate to phosphate along the chain. O is shown red, N blue, P yellow and C grey.

¹Department of Molecular Cell Biology, Samsung Biomedical Research Institute, Sungkyunkwan University School of Medicine, ²Sungkyunkwan Advanced Institute of Nanotechnology, Sungkyunkwan University, Suwon 440-746, Korea. ³Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ⁴Department of Biochemistry, College of Medicine, Chung-Ang University, Seoul 156-756, Korea.

much higher temperature factor, consistent with the weak electron density map around adenine A0.

Base-pair parameters and base-pair step parameters of Z- and B-DNA in this structure are consistent with those observed in standard Z- and B-DNA, respectively^{14,15} (Supplementary Table 3, ref. 16 and Supplementary Discussion, refs 17–20). The single exception is a relatively long rise distance (D_z) of 4.60 Å between the C-2·G2' and C-1·G1' base pairs in the Z conformation (Fig. 3a and Supplementary Table 3), which may be attributable to the fact that the C-1·G1' base pair in the Z conformation has moved from the Z-DNA side to the B-DNA side, in order to stack with T1·A-1' in the B conformation (Figs 3 and 4).

The most striking finding in this structure is that the handedness of the DNA duplex can be completely reversed merely by breaking one base pair and extruding the bases from the duplex. Although in this structure an A and a T are extruded, it is equally possible that the extruded bases could be a G and a C. Base stacking is continuous from B-DNA to Z-DNA through the B-Z junction without any abrupt breaks (Figs 1, 3 and 4); this confirms that base stacking constitutes a major stabilizing factor, even within the B-Z junction. Only a slight destabilization of DNA containing B-Z junctions was observed in thermodynamic investigations of the melting of oligomers containing a B-Z junction¹²; in that study, the transition from B-DNA to the hybrid B/Z DNA structure decreases the free energy of melting by only 0.5 kcal mol⁻¹ relative to the right-handed DNA control. This small energy is not surprising because of the continuity of base stacking through the junction.

The handedness abruptly changes from left to right via a very sharp turn by the phosphate backbone. The junctional angles found

between the three phosphate atoms (P-1, P0 and P1; P-1', P0' and P1') in the B-Z junction are 91.9° and 59.4°, respectively, while the angles between the three adjacent phosphate atoms of standard B- and of Z-DNA are 150° and 110°, respectively. Some adjustments are found for C-1 and T1 and their paired bases. Thus, it would be reasonable to consider that the B-Z junction comprises the three base pairs at the -1, 0 and 1 positions. This is consistent with the previous NMR and Raman spectroscopic studies, which asserted that three base pairs or less may be involved in the B-Z junction^{10,11}.

It was apparent in the first Z-DNA structure¹⁴ that the conversion from the B to Z conformation occurs with the base pairs 'flipping over' and the deoxyguanosine base adopting the *syn* conformation (Supplementary Fig. S5). Seeing the structure of the B-Z junction inspires the formulation of a model for the B-DNA to Z-DNA transition. Torsional strain causes the breaking of a base pair and extrusion of the bases. The extruded bases are free to flip over and reform a base pair, thereby nucleating base-pair breaking and extrusion in the adjacent pair of bases. This process could continue, zipper-like, in two directions, to the limit of the Z-DNA forming sequence, with the junctional bases remaining extruded. Additional negative torsional strain in chromatin could lengthen the Z-DNA segment with continuation of the base-pair disruption, extrusion and reformation. At the junctional sites, slight adjustments rearrange base stacking and minimize backbone strain.

It has been established through structural analysis that base extrusion from the double helix is common in base-modifying enzymes²¹. The similarity of those structures, especially with the conformation of A0, raises the possibility that the extruded bases

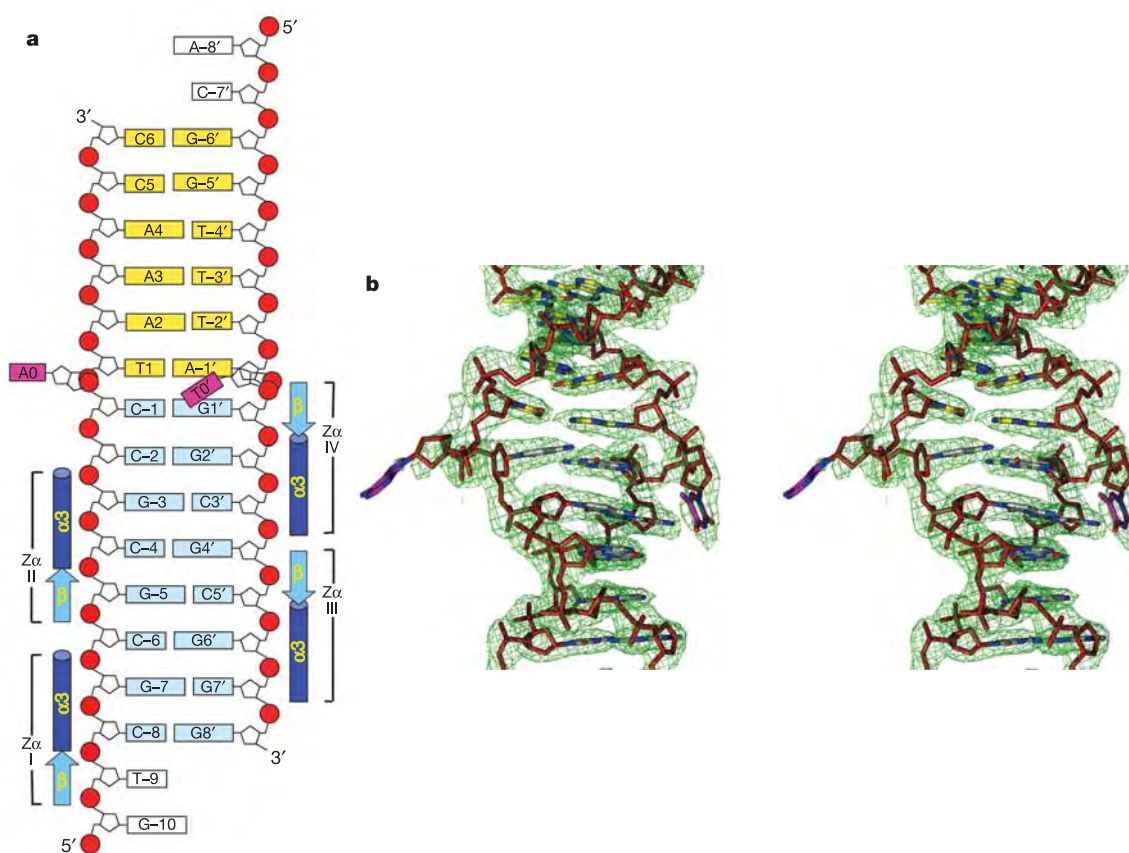


Figure 2 | Organization of the complex and electron density maps. **a**, Schematic diagram of the DNA sequence and protein positions. One strand has numbers with a prime symbol. Negative numbers start at the 5' end and increase towards the extruded junctional bases A0 and T0'. Positions of the DNA recognition helix and β -wing of Z α are indicated.

There are no contacts between the protein and the extruded bases. **b**, Stereo $2F_o - F_c$ omit map of the centre of the molecule with A0 and T0' omitted, calculated using native data at 2.6 Å. The view is from the rear with respect to Fig. 1.

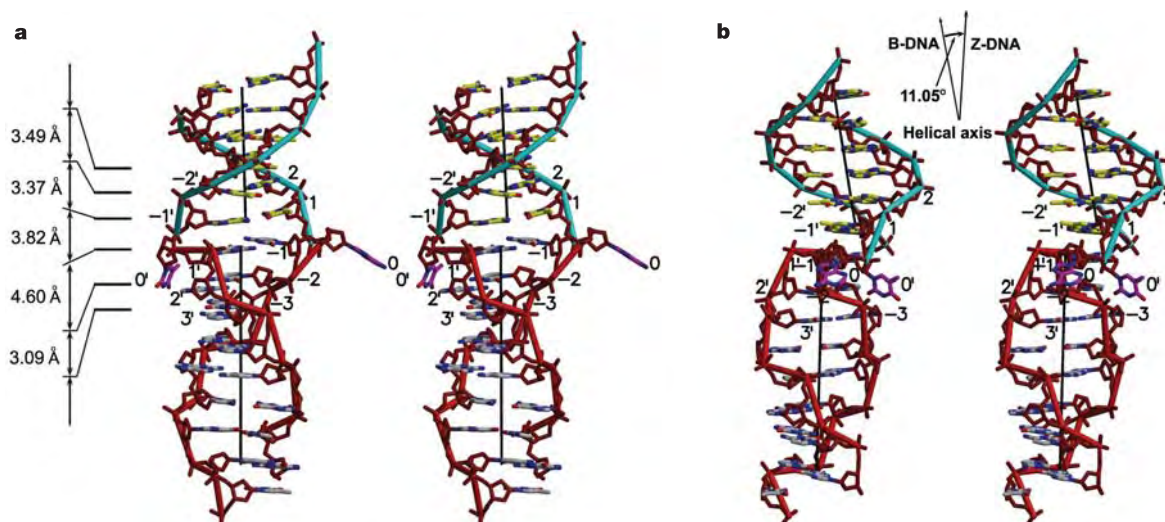


Figure 3 | Skeletal stereo views of DNA. Red and blue lines connect phosphate groups in Z-DNA and B-DNA, respectively. Numbers refer to bases, as shown in Fig. 2. **a**, Front view showing A0 (right) and T0' (left), and the abrupt reversal of helical twist. Distances between stacked bases in the

central region (G-3, C-2, C-1, A0, T1, A2 and A3 on one strand, and T-3', T-2', A-1', T0', G1', G2' and C3' on the other) are indicated. **b**, Side view with A0 in front. The deviation in helical axes between B- and Z-DNA is shown.

might also be recognizable as substrates by some base-modifying enzymes. The A0 base is similar in position to that seen in the crystal structures of extruded bases co-crystallized with HhaI methyltransferase²², human DNA repair protein AGT (O(6)-alkylguanine-DNA alkyltransferase)²³, or bacteriophage T4 endonuclease V (ref. 24) (Supplementary Fig. S6). In addition, the exposed bases may make them susceptible to chemical modification by oxidizing agents *in vivo*.

Our knowledge of the biological role of Z-DNA has been broadened by the discovery of four families of proteins that bind to it, most of which have been characterized by protein-Z-DNA crystal structures. These include the editing enzyme ADAR1 (ref. 5), an interferon-inducible Z-DNA binding protein DLM-1 (ref. 6), a pox virus virulence factor protein E3L (refs 7, 8), and more recently an orthologue of the interferon-induced kinase PKR, which has dsRNA binding domains in mammals but contains Z-DNA binding domains in fish (PKZ)⁹.

There are several observations that point to a role for Z-DNA in transcriptional regulation. The widespread occurrence of sequences favouring Z-DNA formation near the promoter region of genes² places them in a position where they can trap negative supercoiling, which accumulates behind a moving polymerase or is associated with nucleosome unwrapping. It has been shown that transcription of the colony stimulating factor-1 gene requires Z-DNA formation in that region; and since Z-DNA cannot go into nucleosomes, it leaves a region of DNA free to accept transcription factors³. In other examples, such as the constitutive nucleolin gene, promoter sequences forming Z-DNA down regulate expression²⁵. The most striking biological effect of a Z-DNA binding protein is seen in the vaccinia pox virus interferon resistance protein E3L. Vaccinia virus is lethal for mice, but if the Z-DNA binding ability is weakened, pathogenicity is decreased. If Z-DNA binding is lost, the virus is no longer lethal for the mouse⁷. Recent experiments have shown that expressing E3L in HeLa cells transactivates some HeLa genes and

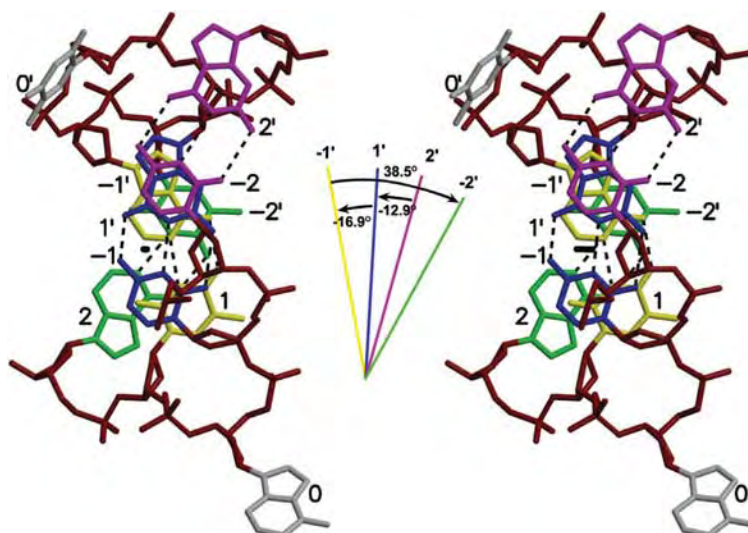


Figure 4 | Stereo skeletal view perpendicular to the bases covering the five central bases. C-2, C-1, A0, T1 and A2 are in one strand, T-2', A-1', T0', G1' and G2' in the other. Successive changes in the orientation of

vectors between deoxyribose atoms C1'-C1'' of the base pairs is shown, illustrating the left-handed rotation in the upper base pairs and the right-handed reversal in the lower base pair.

inhibits apoptosis, both of which are dependent upon Z-DNA binding²⁶.

Every time a section of DNA forms a left-handed Z-DNA segment, two B-Z junctions are formed. The significance of the observed junction is that it requires little structural disruption, preserves helix integrity and, through maximizing base pairing, it minimizes the energetic cost of the junction. This makes it possible to use the Z-DNA conformation more widely and may be related to the widespread occurrence of sequences favouring Z-DNA formation². B-Z junctions are easily formed in biological systems owing to the widespread occurrence of negative supercoiling. These can be further stabilized by Z-DNA binding proteins. The knowledge that these are associated with extrusion of bases at the junction is likely to open many new avenues of research.

METHODS

Crystallization. The protein and DNA duplex in 5 mM HEPES-NaOH pH 7.5 and 10 mM NaCl were mixed to final concentrations of 0.8 mM protein and 0.4 mM DNA (single strand), and crystallized by hanging drop vapour diffusion. Optimized hexagonal crystals were grown in reservoir solutions containing 55–60 mM Na-acetate-HCl pH 4.6, 22–23% MPD and 15–16 mM CaCl₂. For selenium MAD analysis, Lys 179 was mutated to methionine.

Structure determinations. A native data set was collected at the BL41-XU beamline of Spring-8 (Harima, Japan). The selenium-labelled crystal data were collected at the BL-6B beamline of PLS (Pohang, Korea) at –80 °C. Diffraction data were processed using HKL2000²⁷ and SAINT (Bruker), respectively. Selenium sites were found and refined using SOLVE²⁸ and initial phases were modified using RESOLVE²⁹. The Z-DNA binding protein domains (Zα; chain C, PDB code 1QBJ) were fitted into auto-traced Cα models built by RESOLVE. The Z-DNA model (chains C and D, PDB code 1SFU) and B-DNA model (chains A and B, PDB code 1D23) were fitted into modified electron density maps. Refinements used CNS³⁰ and then Refmac³¹. The final model contains four hZα_{ADAR1} peptides and one DNA duplex. Data collection phasing and refinement statistics are summarized in Supplementary Table 1.

CD measurement. The CD spectra were obtained at 25 °C, using an AVIV model 202. Protein was gradually added to 5 μM dsDNA in CD buffer (10 mM HEPES, pH 7.4, 10 mM KF and 0.1 mM EDTA), as indicated by N/P ratios (Supplementary Fig. 3). Spectra were recorded at intervals of 1 nm, averaged over 5 s.

For detailed methods, see Supplementary Information.

Received 6 April; accepted 28 July 2005.

- Rich, A. & Zhang, S. Timeline: Z-DNA: the long road to biological function. *Nature Rev. Genet.* **4**, 566–572 (2003).
- Champ, P. C., Maurice, S., Vargason, J. M., Camp, T. & Ho, P. S. Distributions of Z-DNA and nuclear factor I in human chromosome 22: a model for coupled transcriptional regulation. *Nucleic Acids Res.* **32**, 6501–6510 (2004).
- Liu, R. et al. Regulation of CSF1 promoter by the SWI/SNF-like BAF complex. *Cell* **106**, 309–318 (2001).
- Oh, D. B., Kim, Y. G. & Rich, A. Z-DNA-binding proteins can act as potent effectors of gene expression *in vivo*. *Proc. Natl Acad. Sci. USA* **99**, 16666–16671 (2002).
- Schwartz, T., Rould, M. A., Lowenhaupt, K., Herbert, A. & Rich, A. Crystal structure of the Zα domain of the human editing enzyme ADAR1 bound to left-handed Z-DNA. *Science* **284**, 1841–1845 (1999).
- Schwartz, T., Behlke, J., Lowenhaupt, K., Heinemann, U. & Rich, A. Structure of the DLM-1-Z-DNA complex reveals a conserved family of Z-DNA-binding proteins. *Nature Struct. Biol.* **8**, 761–765 (2001).
- Kim, Y. G. et al. A role for Z-DNA binding in vaccinia virus pathogenesis. *Proc. Natl Acad. Sci. USA* **100**, 6974–6979 (2003).
- Ha, S. C. et al. A poxvirus protein forms a complex with left-handed Z-DNA: crystal structure of a Yatapoxvirus Zα bound to DNA. *Proc. Natl Acad. Sci. USA* **101**, 14367–14372 (2004).
- Rothenburg, S. et al. A PKR-like eukaryotic initiation factor 2α kinase from zebrafish contains Z-DNA binding domains instead of dsRNA binding domains. *Proc. Natl Acad. Sci. USA* **102**, 1602–1607 (2005).
- Sheardy, R. D. & Winkle, S. A. Temperature-dependent CD and NMR studies on a synthetic oligonucleotide containing a B-Z junction at high salt. *Biochemistry* **28**, 720–725 (1989).
- Dai, Z. Y., Thomas, G. A., Everts, E. & Peticolas, W. L. The length of a junction between the B and Z conformations in DNA is three base pairs or less. *Biochemistry* **28**, 6991–6996 (1989).
- Sheardy, R. D., Levine, N., Marotta, S., Suh, D. & Chaires, J. B. A thermodynamic investigation of the melting of B-Z junction forming DNA oligomers. *Biochemistry* **33**, 1385–1391 (1994).
- Herbert, A., Lowenhaupt, K., Spitzner, J. & Rich, A. Chicken double-stranded RNA adenosine deaminase has apparent specificity for Z-DNA. *Proc. Natl Acad. Sci. USA* **92**, 7550–7554 (1995).
- Wang, A. H. et al. Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature* **282**, 680–686 (1979).
- Drew, H. R. et al. Structure of a B-DNA dodecamer: conformation and dynamics. *Proc. Natl Acad. Sci. USA* **78**, 2179–2183 (1981).
- Lu, X. J., Shakked, Z. & Olson, W. K. A-form conformational motifs in ligand-bound DNA structures. *J. Mol. Biol.* **300**, 819–840 (2000).
- Jiang, H., Zacharis, W. & Amirhaeri, S. Potassium permanganate as an *in situ* probe for B-Z and Z-Z junctions. *Nucleic Acids Res.* **19**, 6943–6948 (1991).
- Wolff, S., Wittig, B. & Rich, A. Identification of transcriptionally induced Z-DNA segments in the human c-myc gene. *Biochim. Biophys. Acta* **1264**, 294–302 (1995).
- Mao, C., Sun, W., Shen, Z. & Seeman, N. C. A nanomechanical device based on the B-Z transition of DNA. *Nature* **397**, 144–146 (1991).
- Tashiro, R. & Sugiyama, H. A nanothermometer based on the different π stackings of B- and Z-DNA. *Angew. Chem. Int. Edn Engl.* **42**, 6018–6020 (2003).
- Roberts, R. J. & Cheng, X. Base flipping. *Annu. Rev. Biochem.* **67**, 181–198 (1998).
- Klimasauskas, S., Kumar, S., Roberts, R. J. & Cheng, X. HhaI methyltransferase flips its target base out of the DNA helix. *Cell* **76**, 357–369 (1994).
- Daniels, D. S. et al. DNA binding and nucleotide flipping by the human DNA repair protein AGT. *Nature Struct. Mol. Biol.* **11**, 714–720 (2004).
- Vassilyev, D. G. et al. Atomic model of a pyrimidine dimer excision repair enzyme complexed with a DNA substrate: structural basis for damaged DNA recognition. *Cell* **83**, 773–782 (1995).
- Rothenburg, S., Koch-Nolte, F., Rich, A. & Haag, F. A polymorphic dinucleotide repeat in the rat nucleolin gene forms Z-DNA and inhibits promoter activity. *Proc. Natl Acad. Sci. USA* **98**, 8985–8990 (2001).
- Kwon, J.-A. & Rich, A. Biological function of the vaccinia virus Z-DNA-binding protein E3L: Gene transactivation and antiapoptotic activity in HeLa cells. *Proc. Natl Acad. Sci. USA* **102**, 12759–12764 (2005).
- Otwinowski, Z. & Minor, W. Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).
- Terwilliger, T. C. Map-likelihood phasing. *Acta Crystallogr. D* **57**, 1763–1775 (2001).
- Brunger, A. T. et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Murshudov, G. N., Lebedev, A., Vagin, A. A., Wilson, K. S. & Dodson, E. J. Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr. D* **55**, 247–255 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Schwartz for discussions and technical advice. Use of the Pohang Light Source (PLS) was supported by MOST and POSCO. Use of the Spring-8 is supported by the Japan Synchrotron Radiation Research Institute (JASRI). K.K.K. was supported by a Basic Science Research Grant from the Korea Research Foundation. A.R. acknowledges support from the National Institutes of Health and the Dana and Ellison Medical Foundations.

Author Information Coordinates and structure factor files have been deposited in the Protein Data Bank under accession number 2ACJ. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.K.K. (kkim@med.skku.ac.kr) or Y.-G.K. (ygkimmit@cau.ac.kr).

Structure of *Escherichia coli* RNase E catalytic domain and implications for RNA turnover

Anastasia J. Callaghan¹, Maria Jose Marcaida¹, Jonathan A. Stead², Kenneth J. McDowall², William G. Scott³ & Ben F. Luisi¹

The coordinated regulation of gene expression is required for homeostasis, growth and development in all organisms. Such coordination may be partly achieved at the level of messenger RNA stability¹, in which the targeted destruction of subsets of transcripts generates the potential for cross-regulating metabolic pathways. In *Escherichia coli*, the balance and composition of the transcript population is affected by RNase E, an essential endoribonuclease that not only turns over RNA but also processes certain key RNA precursors^{2–10}. RNase E cleaves RNA internally, but its catalytic power is determined by the 5' terminus of the substrate, even if this lies at a distance from the cutting site^{11–14}. Here we report crystal structures of the catalytic domain of RNase E as trapped allosteric intermediates with RNA substrates. Four subunits of RNase E catalytic domain associate into an interwoven quaternary structure, explaining why the subunit organization is required for catalytic activity. The subdomain encompassing the active site is structurally congruent to a deoxyribonuclease, making an unexpected link in the evolutionary history of RNA and DNA nucleases. The structure explains how the recognition of the 5' terminus of the substrate may trigger catalysis and also sheds light on the question of how RNase E might selectively process, rather than destroy, specific RNA precursors.

E. coli RNase E is one of the largest members of a highly conserved RNase family¹⁵. A catalytic core is contained within the amino-terminal half^{16–18} of RNase E, which in length and sequence closely resembles its paralogue, RNase G^{9,19} (Fig. 1a and Supplementary Fig. S1). The carboxy-terminal half is largely unstructured and poorly conserved²⁰, but in *E. coli* and related proteobacteria this portion organizes and coordinates the activities of the multienzyme RNA degradosome complex²¹.

The crystal structure that we have determined at 2.9 Å is the first for a member of the RNase E/G family, and it reveals that the catalytic domain of RNase E forms a homotetramer with a molecular mass of roughly 260 kDa (Fig. 1). The isolated RNase E protomer is composed of two globular portions, which we refer to as the 'small' and 'large' domains (Fig. 2). The tetramer is organized as a dimer-of-dimers, and the arrangement of the domains within each dimer resembles the blades and handles of an open pair of scissors (Fig. 1b). At the scissor junction point, intra-domain linkers cross and coordinate a zinc ion through a pair of cysteine residues in the conserved CPxCxGxG motif²² (Fig. 1f). The shared-metal site helps to organize the dimer-dimer interface, thus explaining why mutation of the coordinating cysteine residues disrupts the quaternary structure²².

The 'large' domain (residues 1–400) can be divided into subdomains corresponding to established folds (Fig. 2b and Supplementary Fig. S2). One is structurally related to the RNase H

endoribonuclease family, but in RNase E this subdomain seems to have a structural rather than a catalytic function because none of the active-site residues is present. Embedded in the RNase H fold is an S1 domain, which is a widely occurring RNA-binding structural motif (residues 36–118; Fig. 2), and a second fold that participates in the recognition of the 5' terminus of the bound RNA (residues 119–215; Fig. 2). The remaining component of the large domain is similar to the repetitive structural element within the endo-deoxyribonuclease, DNase I (residues 280–400; red in Fig. 2). Their structural congruence has mechanistic implications, as discussed below.

The three RNAs that crystallized together with RNase E contain the targeted sequence 5'-ACAGUAAUUG-3' found at the 5' end of the antisense regulator of ColE1-type plasmid replication, RNA I²³ (Supplementary Fig. S3). The RNAs have 2'-O-methyl groups, which permit binding but prevent degradation in solution studies. The RNA molecule is accommodated in a channel and is contacted on one side by the S1 domain, 5' sensor, DNase I-like and RNase H-like subdomains of one protomer (protomer B in Fig. 3a), and on the other side by residues of the DNase I-like subdomain from the partner protomer (protomer A in Fig. 3a).

Whereas the 10-mer and 13-mer RNAs are bound entirely by one tetramer, the 15-mer RNA is long enough for it to be shared between two different tetramers in the crystal lattice (Fig. 4a), and this RNA extends from the 5' sensing pocket of one tetramer into the binding channel and catalytic site of a symmetry-related tetramer. Two molecules of 15-mer RNA pass each other in antiparallel orientations and form a highly distorted duplex. We suggest that the 15-mer complex might mimic the association of RNase E with RNA substrates in which the sites of primary cleavage are separated from a single-stranded 5' terminus by a structured segment. A model of such a hairpin is shown in Fig. 4b.

The group on the 5' terminus of the RNA substrate can affect the catalytic rate constant of RNase E by orders of magnitude¹² but does not seem to have a marked effect on substrate binding^{24,25}. A 5'-hydroxy group or triphosphate cap impedes activity, whereas the catalytic power is greatest for substrates with a 5'-terminal monophosphate. The RNAs that crystallize together with the catalytic domain of RNase E contain a 5'-terminal monophosphate, and this is accommodated in a pocket formed between the 5' sensor and RNase H-like subdomains (Fig. 3). The terminal phosphate is engaged by a semicircular ring of hydrogen-bonding donors from the side chain and peptide amide of Thr 170 and the guanidino group of Arg 169. The interaction of Arg 169 is consolidated by a hydrogen bond to the peptide backbone of Gly 124 in the neighbouring strand. The terminal base forms a hydrophobic contact with the side chain of Val 128 (Fig. 3c). Replacement of Arg 169 or Val 128 decreases the

¹Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, UK. ²Asbury Centre for Structural Molecular Biology, University of Leeds, Leeds LS2 9JT, UK. ³Department of Chemistry and Biochemistry and the Center for the Molecular Biology of RNA, Sinsheimer Laboratories, University of California at Santa Cruz, Santa Cruz, California 95064, USA.

enhancement of cleavage seen for substrate with 5' monophosphate, and eliminating the hydrogen bond between the terminal phosphate and Thr 170 by replacement with Val inactivates the enzyme (J.A.S., B.F.L. and K.J.M., unpublished observation). Arg 373 from the RNase H-like subdomain bridges the phosphate at positions 2 and 5, to help dock the 5' terminus and redirect the RNA towards the catalytic site. Replacement of Arg 373 by either Ala or Asp decreases catalytic activity tenfold, with only a modest effect on RNA-binding affinity (Supplementary Information).

We identified a sequestered magnesium ion, presumably hydrated, that is in proximity to the RNA phosphate backbone (Fig. 3b). This metal is coordinated by the carboxylates of Asp 303 and Asp 346, which are conserved throughout the RNase E/G family (Supplementary Information). The catalytic mechanism of RNase E probably involves the in-line nucleophilic attack of the scissile phosphate by an activated hydroxyl group generated from a water molecule that is coordinated to the magnesium. It can be envisaged that the carboxylates of Asp 303 and Asp 346 might act as general bases to activate the attacking water. The conservative replacement of either of these residues with the isosteric and polar Asn decreased catalytic

activity at least 25-fold while having little effect on RNA binding. The neighbouring Asn 305 supports the orientation of Asp 303 through hydrogen bonding, and the elimination of this interaction by the replacement of Asn 305 with either Asp or Leu caused a decrease in activity (Supplementary Information, Table S1). Asp 303, Asp 346 and Asn 305 are on the surface of the subdomain of RNase E that resembles DNase I, and the catalytic sites of these two different enzymes roughly coincide (Fig. 2c, d).

The first base after the cleavage site is partitioned into a hydrophobic pocket on the surface of the S1 domain (Fig. 3b). The pocket is formed on one side by Phe 67, with which the base's aromatic face makes an alternative stacking contact, and on the other side by the aliphatic portion of the Lys 112 side chain. The latter interaction orients the lysine so that its terminal amino group hydrogen bonds with the scissile phosphate and presumably supports the charge built up in the catalytic transition state. Phe 57, also from the S1 domain, helps to consolidate this interaction with the RNA by supporting Phe 67. We find that the individual replacement of any of these residues by Ala decreased activity at least 50-fold; moreover, this does not seem to be due to a marked decrease in affinity for the substrate

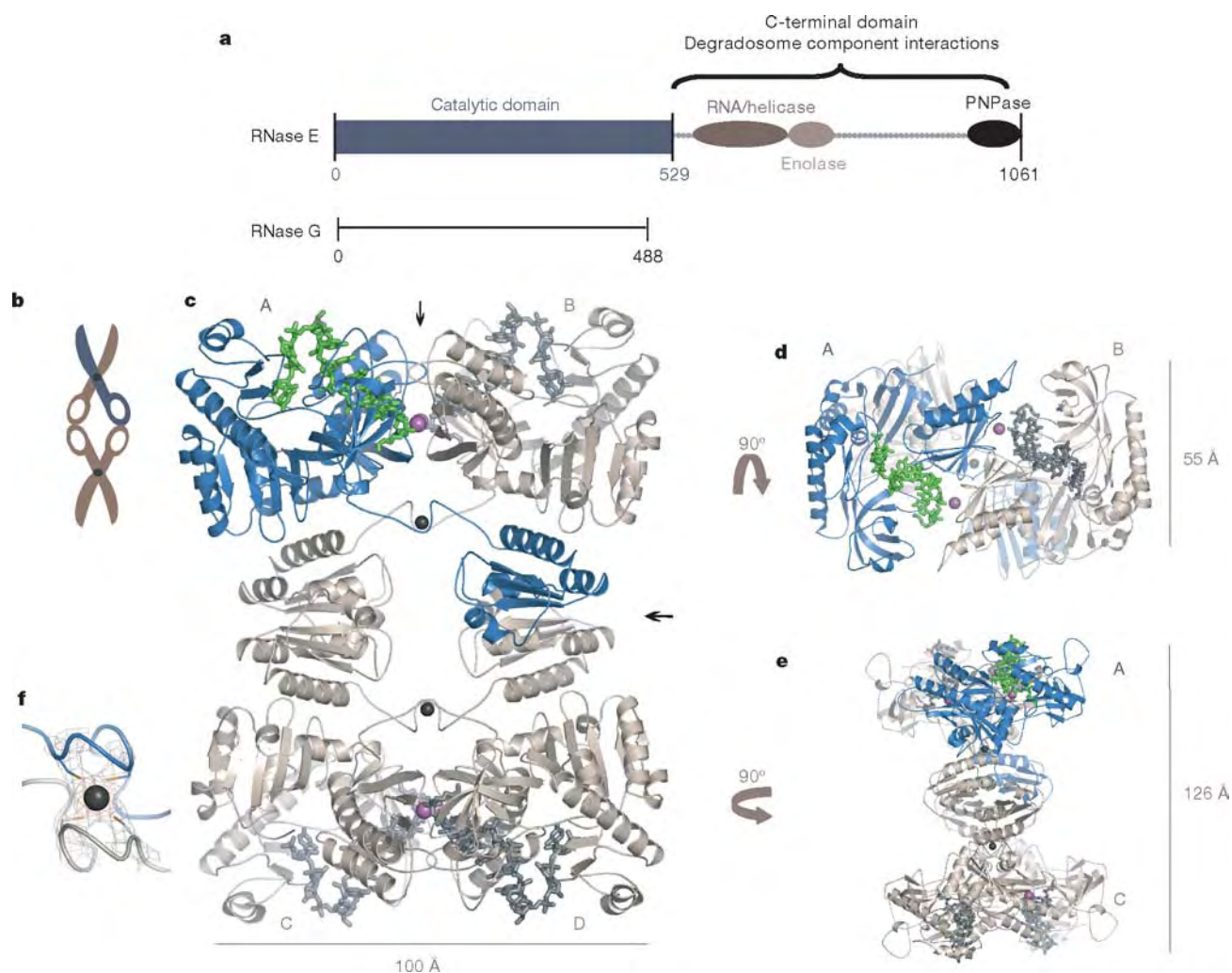


Figure 1 | The structure of RNase E catalytic domain. **a**, Schematic representation of the primary structure of RNase E and paralogue RNase G. The crystallized N-terminal catalytic domain is represented by the blue box. The degradosome-scaffolding C-terminal domain (grey line) has binding regions for RNA/helicase (brown), enolase (beige) and PNPase (black). **b**, Scissor representation of the RNase E tetramer. **c**, The homotetramer, showing protomer A (blue) with bound RNA (green), and the remaining

three protomers B, C and D (beige) with bound RNA (grey). The principal dimers are the pairs A/B and C/D. The zinc and magnesium ions are shown as dark grey and magenta spheres, respectively. **d**, **e**, Homotetramer as viewed from above (**d**) and rotated through 90° (**e**). **f**, The zinc-link coordination site at the scissor junction, formed by two cysteine residues from protomer A (blue) and two from protomer B (beige). Electron density ($2mF_o - DF_c$ coefficients) is shown at 8σ (orange) and 3σ (grey).

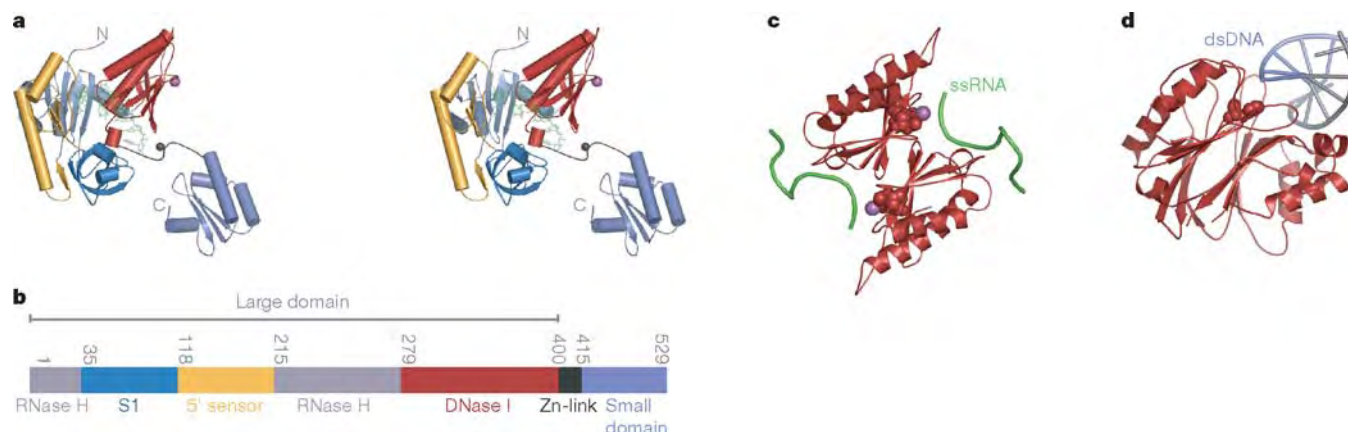


Figure 2 | RNase E isolated protomer-RNA complex. **a**, Stereoscopic view of the isolated protomer-RNA complex, showing the subdomains S1 (blue), RNase H (grey), 5'-sensor (gold), the small domain (violet) and DNase I (red). The single-stranded RNA (green) is semi-transparent for clarity. Zinc and magnesium ions are grey and magenta, respectively. **b**, Linear

representation of the subdomain boundaries. **c**, **d**, Dimerization of the DNase I subdomains in the principal dimer (**c**) and the similar fold of DNase I structure (**d**). Shown are the active site residues (space-filling representation), single-stranded RNA (ssRNA, green) and the double-stranded DNA (dsRNA, grey).

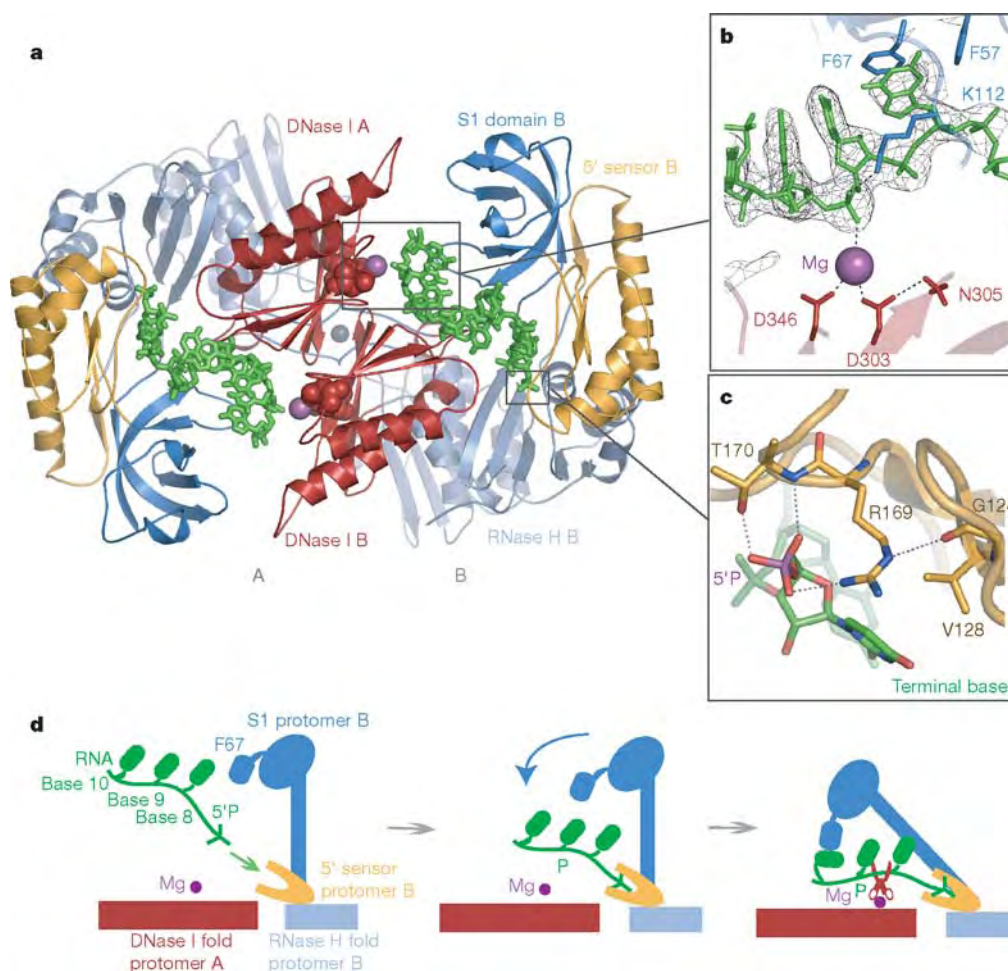


Figure 3 | The RNA-binding channel. **a**, The principal dimer viewed along the twofold symmetry axis. The RNA is contacted by the S1, 5'-sensor, RNase H-like and different parts of the DNase I-like subdomains of protomers A and B. **b**, The proposed catalytic site of RNase E. The magnesium ion (magenta) interacts with a phosphate of the RNA (green), and Asp 346 and Asp 303 of protomer A (red). The $2F_o - F_c$ -style electron density map (black) is at 2.5σ . **c**, The 5' sensing pocket. RNA (green) and 5' sensor residues (gold) are shown. The terminal phosphate of the RNA

(purple) is engaged by a semicircular ring of hydrogen-bond donors, Thr 170 and Arg 169. **d**, A 'mouse-trap' model for communication between the 5' sensing pocket and the site of catalysis. On binding the 5' phosphate, the 5' sensor (gold) induces the S1 domain (blue) to clamp down on the RNA (green) near the scissile phosphate (green P). This orients the RNA for cleavage at the Mg ion (magenta) in the DNase I domain (red). The proposed movement of the S1 domain is exaggerated for illustrative purposes.

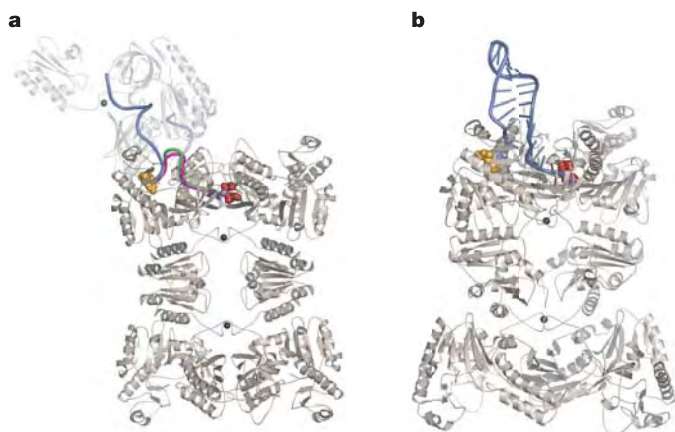


Figure 4 | RNA recognition by RNase E catalytic domain. **a**, An overlay of RNase E tetramer in complex with bound 10-mer (magenta), 13-mer (green) and 15-mer (violet) RNA. Catalytic residues (red) and 5' sensor pocket (gold) in space-filling representation. The 15-mer RNA extends away from the principal dimer into a symmetry-related tetramer (semi-transparent beige) and, reciprocally, from the symmetry-related tetramer (semi-transparent violet) back into the principal dimer (beige). Only one active site with RNA bound is shown for simplicity. **b**, RNase E bound to an RNA hairpin model (violet) extrapolated from the RNase E/15-mer RNA complex crystal structure.

(Supplementary Table S1). Perhaps the sandwiching of the base between the Phe67 and Lys112 distorts the RNA to favour the transition state.

In all the complexes, the RNA follows the same stacking arrangement over the surface of the RNA-binding channel, and only one of the bases forms a hydrogen bond with the protein; Lys106 contacts the exocyclic oxygen of the base at position -1 with regard to the scissile phosphate. This direct contact might explain the preference for G or U at this position²⁶. Aside from that contact, there is no sequence recognition as such, so it seems that the preference of RNase E for A + U-rich substrates^{27–29} arises mainly through the recognition of the RNA conformation.

Our crystal structure shows that the catalytic site and the 5' sensing site are physically separated, yet they must somehow communicate so that the status of the 5' end of the substrate is relayed to the active site. It seems most likely that the communication is mediated through allosteric change in the protein. In support of this proposal, we observe that the binding of RNA induces compaction of the protein using neutron solution scattering with contrast variation (A.J.C., J.G. Grossmann, M.J.M. and B.F.L., unpublished observation). On the basis of the crystal structure and the solution experiments, we propose a 'mouse-trap'-like mechanism that explains how 5' sensing is linked to catalysis through an allosteric change in the protein (Fig. 3d). The engagement of the 5' monophosphate of the RNA organizes the 5' sensor pocket, which in turn interacts with the S1 domain and causes it to clamp down on the RNA downstream, thus orienting the phosphate backbone to favour the in-line attack on the scissile phosphate by the magnesium-coordinated, activated hydroxyl group. Thus, catalysis proceeds through classical induced-fit.

The question naturally arises as to why RNase E does not indiscriminately cleave all RNA with a 5' monophosphate, including the structured substrates that it processes, such as ribosomal RNA⁹ and RNase P precursors⁷. The answer may lie in the mouse-trap mechanism, which requires that the 5' end is engaged in the sensing site and that the portion to be cleaved is simultaneously engaged at the active site: structured RNA segments may not be able to satisfy all of these requirements simultaneously.

RNase E/G proteins are important in global gene regulation and cellular homeostasis through their preferential degradation of transcripts and their processing of key RNA species. Their homologues in

plants are also implicated in the general control of gene expression. The crystal structure of *E. coli* RNase E opens the possibility of manipulating this control, either for therapeutic intervention or for the purpose of examining the complex interaction of different components in dynamic cellular systems.

METHODS

Structure determination. The catalytic domain of RNase E and a selenomethionine derivative (residues 1–529 with an N-terminal hexahistidyl affinity tag) were prepared by the method described previously^{24,30}. For the crystallization trials, the purified protein was mixed in the ratio 1:1.2 with HPLC-purified synthetic RNA containing a 5'-monophosphate and 2'-O-methyl groups and concentrated to 10–15 mg ml⁻¹. Crystals of the RNase E catalytic-domain–RNA complex appeared after 2–4 weeks in 5–20% w/v poly(ethylene glycol) 8000, 0.1 M Tris-HCl pH 7.5–8.0, and 10–50 mM magnesium formate at 20 °C, and in all cases the complex had a hexagonal prismatic habit. Crystals were cryo-protected by serial transfer into crystallization buffer containing 5–30% v/v ethylene glycol, and then frozen by plunging into liquid nitrogen. Diffraction data were collected at 100 K at stations ID 14-4, 23-1 and 29-1 (European Synchrotron Radiation Facility, Grenoble); station 14.2 (Synchrotron Radiation Source, Daresbury); and Howard Hughes Medical Institute station 8.8.2 (Advanced Light Source, Berkeley). The structure was solved by using multiple-wavelength anomalous scattering and multiple isomorphous replacement, as detailed in Supplementary Information. Structural figures were made using PYMOL (DeLano Scientific) and NUCCYL (<http://www.biosci.ki.se/groups/ljo/software/nucyl.html>).

Site-directed mutagenesis and RNase E assays. The method of site-directed mutagenesis of RNase E catalytic domain and expression of the mutant polypeptides in *E. coli* were described previously²². The sequences of the pairs of primers used to introduce the specific mutations are provided in Supplementary Table S3. Each recombinant polypeptide was purified with Ni²⁺-nitrilotriacetate spin columns. The eluate was dialysed against 20 mM Tris-HCl pH 7.9, 0.5 M NaCl, 10 mM dithiothreitol, 10 mM MgCl₂, 0.5 mM EDTA, 5% (v/v) glycerol, divided into aliquots and stored frozen at -70 °C. Protein concentrations were determined with a Bradford protein assay (Bio-Rad), and confirmed by SDS-PAGE.

The cleavage of RNA was assayed with a continuous system similar to that described previously²⁵. The 5' monophosphorylated substrate used here had Dabcyl-dT (D-dT) at position 5 and fluorescein (Fl) at the 3' end. The sequence was 5'-GGGA(D-dT)CAGUAUUU-Fl, which is based on the RNase E cleavage site at the 5' end of RNA I²³. The reaction conditions were those optimized previously²⁴. The concentrations of substrate and enzyme were 500 nM and 8 nM, respectively. Mutants with greatly decreased activity were also assayed at a concentration of 40 nM. The increase in fluorescence was measured with a Perkin Elmer LS50B fluorimeter at 37 °C. The excitation and emission wavelengths were 488 and 523 nm, respectively. Each mutant was assayed in triplicate. The system was calibrated with substrates cleaved to completion by RNase E and RNase A.

For the electrophoretic mobility-shift assays, a 5' monophosphorylated substrate with sequence 5'-ACAGUAUUUG-3', with protective 2'-O-methyl groups at all positions and fluorescein at the 3' end, was used²². The optimized cleavage conditions²⁴ were used to assay RNA binding. Binding was assayed by determining the amount of unbound substrate remaining in each reaction relative to a 5'-fluorescein-labelled pentadeoxyribonucleotide (5'-GACTA-3') that served as an internal control.

Received 3 March; accepted 26 July 2005.

1. Grunberg-Manago, M. Messenger RNA stability and its role in control of gene expression in bacteria and phages. *Annu. Rev. Genet.* **33**, 193–227 (1999).
2. Kuwano, M. et al. Gene affecting longevity of messenger RNA: a mutant of *Escherichia coli* with altered mRNA stability. *Mol. Gen. Genet.* **154**, 279–285 (1977).
3. Ono, M. & Kuwano, M. A conditional lethal mutation in an *Escherichia coli* strain with a longer chemical half life of mRNA. *J. Mol. Biol.* **129**, 343–357 (1979).
4. Bernstein, J. A., Lin, P.-H., Cohen, S. N. & Lin-Chao, S. Global analysis of *Escherichia coli* RNA degradosome function using DNA microarrays. *Proc. Natl Acad. Sci. USA* **101**, 2758–2763 (2004).
5. Mudd, E. A. & Higgins, C. F. *Escherichia coli* endoribonuclease RNase E: autoregulation of expression and site-specific cleavage of mRNA. *Mol. Microbiol.* **9**, 557–568 (1993).
6. Apirion, D. & Lassar, A. B. A conditional lethal mutant of *Escherichia coli* which affects the processing of ribosomal RNA. *J. Biol. Chem.* **253**, 1738–1742 (1978).

7. Lundberg, U. & Altman, S. Processing of the precursor to the catalytic RNA subunit of RNase P from *Escherichia coli*. *RNA* **1**, 327–334 (1995).
8. Ow, M. C. & Kushner, S. R. Initiation of tRNA maturation by RNase E is essential for cell viability in *Escherichia coli*. *Genes Dev.* **16**, 1102–1115 (2002).
9. Li, Z., Pandit, S. & Deutscher, M. P. RNase G (CafA protein) and RNase E are both required for the 5' maturation of 16S ribosomal RNA. *EMBO J.* **18**, 2878–2885 (1999).
10. Kim, K.-S. & Lee, Y. Regulation of 6S RNA biogenesis by switching utilization of both sigma factors and endoribonucleases. *Nucleic Acids Res.* **32**, 6057–6068 (2004).
11. Bouvet, P. & Belasco, J. G. Control of RNase E-mediated RNA degradation by 5'-terminal base pairing in *E. coli*. *Nature* **360**, 488–491 (1992).
12. Mackie, G. A. Ribonuclease E is a 5'-end-dependent endonuclease. *Nature* **395**, 720–723 (1998).
13. Mackie, G. A. Stabilization of circular rpsT mRNA demonstrates the 5' end dependence of RNase E action *in vivo*. *J. Biol. Chem.* **275**, 25069–25072 (2000).
14. Coburn, G. A. & Mackie, G. A. Degradation of mRNA in *Escherichia coli*: an old problem with some new twists. *Prog. Nucleic Acid Res.* **62**, 55–108 (1999).
15. Avarind, L. & Koonin, E. V. A natural classification of ribonucleases. *Methods Enzymol.* **341**, 3–28 (2001).
16. McDowall, K. J. & Cohen, S. N. The N-terminal domain of the *rne* gene product has RNase E activity and is non-overlapping with the arginine-rich RNA binding site. *J. Mol. Biol.* **255**, 349–355 (1996).
17. Kido, M. *et al.* RNase E polypeptides lacking a carboxyl-terminal half suppress a *mukB* mutation in *Escherichia coli*. *J. Bacteriol.* **178**, 3917–3925 (1996).
18. McDowall, K. J., Hernandez, R. G., Lin-Chao, S. & Cohen, S. N. The *ams-1* and *rne-3071* temperature-sensitive mutations in the *ams* gene are in close proximity to each other and cause substitutions within a domain that resembles a product of the *Escherichia coli mre* locus. *J. Bacteriol.* **175**, 4245–4249 (1993).
19. Wachi, M., Umitsuki, G., Shimizu, M., Takada, A. & Nagai, K. *Escherichia coli cafA* gene encodes a novel RNase, designated as RNase G, involved in processing of the 5' end of 16S rRNA. *Biochem. Biophys. Res. Commun.* **259**, 483–488 (1999).
20. Callaghan, A. J. *et al.* Studies of the RNA degradosome-organising domain of the *Escherichia coli* RNase E. *J. Mol. Biol.* **340**, 965–979 (2004).
21. Carpousis, A. J. The *Escherichia coli* RNA degradosome: structure, function and relationship in other ribonucleolytic multienzyme complexes. *Biochem. Soc. Trans.* **30**, 150–155 (2002).
22. Callaghan, A. J. *et al.* The 'Zn-link': a metal-sharing interface that organizes the quaternary structure and catalytic site of the endoribonuclease, RNase E. *Biochemistry* **44**, 4667–4675 (2005).
23. Lin-Chao, S. & Cohen, S. N. The rate of processing and degradation of antisense RNA I regulates the replication of ColEI-type plasmids *in vivo*. *Cell* **65**, 1233–1242 (1991).
24. Redko, Y. *et al.* Determination of the catalytic parameters of the N-terminal half of *E. coli* ribonuclease E and the identification of critical functional groups in RNA substrates. *J. Biol. Chem.* **278**, 44001–44008 (2003).
25. Jiang, X. & Belasco, J. G. Catalytic activation of multimeric RNase E and RNase G by 5'-monophosphorylated RNA. *Proc. Natl Acad. Sci. USA* **101**, 9211–9216 (2004).
26. Kaberdin, V. R. Probing the substrate specificity of *Escherichia coli* RNase E using a novel oligonucleotide-based assay. *Nucleic Acids Res.* **31**, 4710–4716 (2003).
27. Cohen, S. N. & McDowall, K. J. RNase E: still a wonderfully mysterious enzyme. *Mol. Microbiol.* **23**, 1099–1106 (1997).
28. McDowall, K. J., Lin-Chao, S. & Cohen, S. N. A + U content rather than a particular nucleotide order determines the specificity of RNase E cleavage. *J. Biol. Chem.* **269**, 10790–10796 (1994).
29. Lin-Chao, S., Wong, T. T., McDowall, K. J. & Cohen, S. N. Effects of nucleotide sequence on the specificity of *rne*-dependent and RNase E-mediated cleavages of RNA I encoded by the pBR322 plasmid. *J. Biol. Chem.* **269**, 10797–10803 (1994).
30. Callaghan, A. J. *et al.* Quaternary structure and catalytic activity of the *Escherichia coli* ribonuclease E amino-terminal catalytic domain. *Biochemistry* **42**, 13848–13855 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the staff of the Daresbury Synchrotron Radiation Source, the European Synchrotron Radiation Facility and the Advanced Light Source synchrotrons for the use of facilities. We thank A. Murzin for pointing out the similarity with DNase I, and A. and T. Andreeva for helpful discussions about the structural deconvolution. We thank V. Chandran, K. Nagai, A. J. Carpousis and M. Symmons for advice. We thank C. Hill for synthesis of the protected RNA and L. Packman for protein and peptide analyses. This work was supported by the Wellcome Trust.

Author Information The coordinates and structure factors have been deposited in the Protein Data Bank (PDB ID 2BX2, R2BX2SF and 2COB). Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to B.F.L. (ben@cryst.bioc.cam.ac.uk).

More than than gene expression

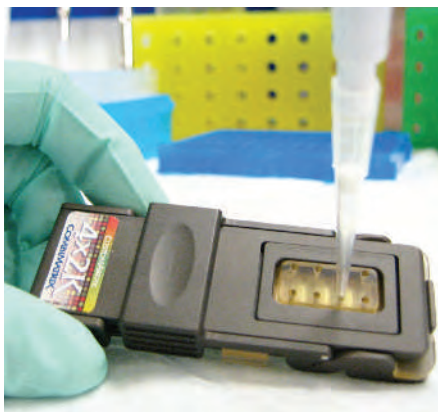
DNA microarrays are diversifying in new directions, including *in vitro* diagnostics. Diane Gershon takes a look at what's around the corner for microarray applications.

COMBIMATRIX

It's now ten years or so since Affymetrix of Santa Clara, California, first began selling commercial DNA microarrays. In that time, complete or partial sequences of numerous genomes, including human, have been deposited in the public databases. Coupled with recent improvements in microarray technology, this development is propelling microarrays into areas of application that extend well beyond gene-expression analysis.

It is also now generally accepted that new insights will not be gained simply by acquiring more and more gene-expression data, and that it is no longer sufficient to focus on the 25,000 or so protein-coding genes that make up roughly 2% of the human genome. Exploring the role and diversity of non-coding RNAs is equally important.

Ten years on, Affymetrix still dominates the high-density DNA microarray market. But new players have entered more recently, hoping to carve out a niche in new and emerging areas such as splice-variant or microRNA analysis (see 'It's a small world'). Others are differentiating themselves by leveraging their more flexible methods of array synthesis and offering custom-made microarrays for the



Flexibility at low density in CombiMatrix's 4 × 2K CustomArrays.

human genome and a whole host of model organisms.

As a general rule of thumb, there are three key things to look out for when it comes to assessing DNA microarray platforms: feature density per array, flexibility (how easy it is to change the array content and design), and sensitivity. CombiMatrix of Mukilteo, Washington, an operating group of Acacia Research

Corporation, concentrates on designing and manufacturing relatively low-density custom microarrays containing up-to-date content. Last month the company began offering its 4 × 2K customized microarrays for US\$99, produced using its semiconductor-based *in situ* oligonucleotide synthesis method. The 4 × 2K CustomArrays contain four arrays of 2,240 oligonucleotide features per slide with a 44-μm feature size. CombiMatrix plans to offer its first high-density array early in 2006, which will have 90,000 features and a 25-μm feature size.

"We're custom, quick, flexible and cost-effective," says Michael Tognotti, vice-president of sales and marketing at CombiMatrix. There are no minimum order requirements for these arrays, which Tognotti says can be re-used up to three times, further driving down costs for researchers.

Tiling technology

But companies such as Affymetrix see less of a need for customization now that it can put whole genomes on a chip. "Affymetrix, as a result of continually shrinking the feature size, has made it practical to do experiments where

IT'S A SMALL WORLD

MicroRNAs (miRNAs) represent a new market opportunity for microarray companies (see *Nature* 435, 991-996; 2005). "MicroRNA is as hot as it could possibly be right now," says Scott Cole, head of genomics marketing at Agilent Technologies of Palo Alto, California, which expects to launch specific miRNA array-based products in 2006.

Exiqon of Vedbaek, Denmark, is now focused on developing miRNA detection products based on its locked nucleic acid (LNA) technology. "An LNA oligo will bind tighter to a DNA than a DNA itself will, and even more so to an RNA," says Mikkel Nørholm, senior research scientist at Exiqon. "Because the affinity is higher you can use a shorter probe," he says. The company offers ready-to-spot oligonucleotide capture probes based on the latest information from the miRBASE sequence database (<http://microrna.sanger.ac.uk>).

A pre-spotted array for microRNA detection in human and mouse will be available soon.

Last month, Ambion of Austin, Texas, inked a deal with Rosetta Genomics of Rehovot, Israel, to access Rosetta's proprietary miRNA sequence database, which will add to its current range of microarrays targeting human, mouse and rat miRNAs from miRBASE. Ambion's mirVana miRNA Bioarrays are manufactured by GE Healthcare on its CodeLink platform, which uses a three-dimensional (3D) gel matrix to lift the probes off the slide surface. This is designed to maximize interaction between probe and target and enable the detection of low-abundance miRNAs.

miRNA arrays from LC Sciences of Houston, Texas, are currently available on a service basis. The addressable microfluidics chip contains almost 4,000 picolitre-



Scott Cole: "miRNA is as hot as it possibly could be".

volume 3D chambers and stems from work at the University of Houston and the University of Michigan. At the heart of the system is the μParaflo microfluidic technology, which enables fast *in situ* parallel synthesis of large numbers of different oligos at high

yield. In addition to probes based on sequences in miRBASE, researchers can add up to 100 custom sequences at no extra charge. miRNA arrays and services are also available from GenoSensor of Tempe, Arizona, and Paradigm Array Labs of Research Triangle Park, North Carolina, a service unit of Icoria.

Kreatech Biotechnology of Amsterdam, The Netherlands, plans to extend its Universal Linkage System (ULS) direct labelling technology to miRNA research by offering a labelling kit optimized for small RNAs. According to Brent Keller, Kreatech's vice-president of commercial applications, ULS provides a fast and simple alternative to enzymatic labelling methods, which can be subject to 3'-end bias. "Our chemistry is independent of fragment length," says Keller, "making it ideal for miRNA applications."

D.G.

AGILENT



Looking close: Affymetrix's high-density tiling arrays interrogate the genome in detail.

we look across the entire genome without bias," says John Blume, vice-president of RNA products at Affymetrix.

The development of tiling arrays marked a departure from the way DNA microarrays had traditionally been designed. Rather than containing probes just for known or suspected functional regions, tiling arrays can be used to interrogate, or 'tile' across, the genome in an unbiased fashion, by means of probes spaced at regular intervals along the genome.

Tiling microarrays can be used for the genome-wide study of gene regulation, specifically to map sites of transcription-factor binding, chromatin modification, DNA methylation and chromosomal origins of

replication. This month, Affymetrix will launch a set of seven arrays, each with 6.5 million probes and a 5- μ m feature size, to tile across the entire human genome at 35-base-pair (bp) resolution. Tiling arrays for several model organisms, including *Arabidopsis*, *Drosophila*, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, are planned.

"Affymetrix will continue to innovate with respect to feature size. We don't think 5 μ m is the end," says Blume.

The highest-density arrays from Agilent Technologies of Palo Alto, California, have 44,000 features. The company is about to launch its next-generation ink-jet printing platform for the non-contact *in situ* synthesis of 60-mer oligonucleotide probes on glass slides. With the new instrument, which will not be sold commercially, "we think that we can get about an order of magnitude higher density in the span of about two years," says Scott Cole, head of genomics marketing at Agilent.

Scanning to size

The ever-shrinking feature size on high-density microarrays must also be coupled with improvements in scanning technology. This summer Affymetrix launched the new GeneChip Scanner 3000 7G with a scanning resolution in the sub-micrometre pixelation range, and which can scan features ranging in size from 2.5 to 0.51 μ m. The scanner supports all high-resolution GeneChip products for tiling, all-exon and single-nucleotide polymorphism genotyping research.

For those on a budget, TeleChem International offers the 10- μ m resolution SpotLight microarray scanner. Rather than using lasers and photomultiplier tubes, as in most traditional scanners, the SpotLight uses 'cool' excitation technology that is optimized for two-colour detection (Cy3 and Cy5). Optional filter sets are available for fluorescein, rhodamine, allophycocyanin, ethidium bromide, the Alexa dyes and several green fluorescent protein variants. Scanning solutions are also available from established scanner manufacturers including arrayWoRxe from Applied Precision of Issaquah, Washington, GenePix from Molecular Devices of Sunnyvale, California and ProScan Array from Perkin Elmer.

CombiMatrix is developing an electrochemical detection (ECD) method to read its semiconductor-based microelectrode arrays, as an alternative to fluorescence. The company hopes to develop a smaller, less expensive detection system that will provide better performance than fluorescence-based systems — the only method so far to be commercially successful. A prototype is in beta testing and a commercial product is expected next year. CombiMatrix hopes to corner the molecular diagnostics market with a handheld version.

Options are also available for researchers who want to print their own microarrays (see 'On the hardware front', page 1198).

Comparative genomic hybridization

For the best part of a decade, the microarray market has been dominated by gene expression. "But gene expression is not the most

MICROARRAYS MOVE DOWNSTREAM

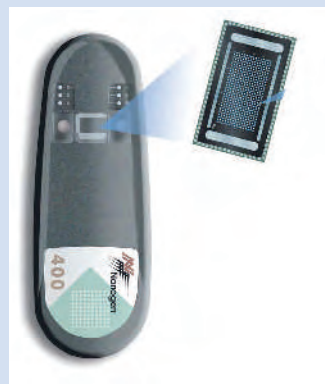
Array-based methods are now being applied in a diagnostic setting and offer the possibility of analysing multiple analytes (and even multiple samples) in a highly parallel and high-throughput manner. In diagnosis, however, the tolerance for error is very low: it is important to get the right answer not just occasionally but all of the time, under stringent conditions and with varying sample quality. Nevertheless, important milestones have been achieved in the past year as array-based diagnostic tests begin to trickle onto the market.

Agendia of Amsterdam, The Netherlands, offers MammaPrint, a diagnostic test for assessing the risk of breast cancer progression. The test predicts the future development of metastases on the basis of a 70-gene expression signature discovered by scientists at the Netherlands Cancer Institute and

Antoni van Leeuwenhoek Hospital in Amsterdam. Its use should improve the quality of treatment decisions for cancer patients by distinguishing those individuals who would, and who would not, benefit from further treatment.

In May, Tm Bioscience of Toronto, Canada, received FDA clearance for its Tag-It cystic fibrosis (CF) kit to be used as an *in vitro* device for diagnostic use. The test, which identifies a group of mutations in the CF transmembrane conductance regulator (CFTR) gene, is based on the bead-based xMAP multiplexing technology developed by Luminex of Austin, Texas. xMAP allows multiplexing of up to 100 assays within a single sample. In addition to human genetic disorders, Tm Bioscience is developing products aimed at pharmacogenetics and infectious diseases.

Last year, Affymetrix of Santa



The NanoChip electronic microarray goes diagnostic.

Clara, California, received clearance by regulatory authorities in the United States and Europe for the *in vitro* diagnostic use of its GeneChip System 3000Dx. Affymetrix's strategy for developing new array-based diagnostic tests on its GeneChip platform is to work alongside partners from the

diagnostics industry, such as Roche Diagnostics of Basel, Switzerland, bioMérieux of Marcy l'Etoile, France, and Veridex of Warren, New Jersey. "We are a technology-driven company," says Robert Lipshutz, senior vice-president of molecular diagnostics and emerging markets at Affymetrix. "We did not have a large regulatory or commercial presence in the diagnostics industry."

The first microarray product to be launched through the partnership with Roche is the AmpliChip CYP450 test, which identifies variations in two drug-metabolism genes that can affect the rate at which an individual metabolizes many common drugs.

Other companies are taking a different approach. CombiMatrix of Mukilteo, Washington, formed a diagnostic subsidiary, CombiMatrix Molecular Diagnostics, in May, now led by

demanding thing you can do with a microarray," says Emile Nuwaysir, vice-president of business development at NimbleGen of Madison, Wisconsin. Other microarray applications are now coming to the fore "that even have the potential to eclipse gene expression," he says. One such area is array-based comparative genomic hybridization (CGH), which can provide a robust and accurate platform for routinely mapping chromosomal imbalances at exon-level resolution.

Several companies now offer microarray-based products and services for identifying and characterizing gains and losses of genomic regions that lead to copy-number changes in genes and regulatory regions. Such changes are thought to underlie diseases such as cancer and many congenital disorders. "To be able to see in the full complexity of the human genome that small change in signal is very demanding, and commercial arrays weren't capable of doing it five years ago," says Nuwaysir.

Unlike some array-based CGH products that use a gene-centric design, NimbleGen's human whole-genome array CGH platform is a single array with 385,000 probes tiled throughout the human genome every 6,000 bp. The company's customized fine-tiling CGH arrays contain the same number of probes as the whole-genome CGH array but tiled at higher resolution in chromosomal regions selected by the researcher. Here the probes can be spaced as densely as every 10 bp.

"We're the only company that combines the benefits of high-density, long oligo arrays and

the ability to change the design whenever you want," says Nuwaysir. NimbleGen also hopes that an isothermal probe-selection strategy will set it apart from the competition. Rather than fix the oligo length, as most companies do, NimbleGen fixes their melting temperatures at 76 °C across the entire set. As such, says Nuwaysir, the probes are optimized to perform equivalently in all genomic regions, including AT- and GC-rich regions.

Agilent introduced an updated version of its microarray-based CGH platform in August. The company already offers a microarray for the genome-wide scanning of chromosomal gains and losses on a single chip. Now, researchers will be able to tap into Agilent's database of approximately 4 million validated CGH probes and use its web-based DNA microarray design tool, eArray, to create custom CGH microarrays for high-resolution tiling of regions of interest. The company also recently launched a single-slide CGH microarray for the genome-wide profiling of DNA copy-number changes in mouse.

Location, location, location

A second expanding research area now being adapted to a microarray format is chromatin immunoprecipitation (chIP) analysis. Also referred to as 'location analysis', chIP followed by microarray analysis of the immunoprecipitated genomic DNA, the so-called chIP-chip methodology, can be used to identify the binding sites on chromosomes for DNA-binding proteins such as histones, polymerases and transcription factors, as well as studying DNA

modification and chromatin-remodelling.

NimbleGen offers a microarray service for chIP-chip analysis. Researchers can either select from off-the-shelf array designs, which include a whole-genome survey set with 390,000 features per array and two array designs aimed at known promoter regions, or can create their own customized tiling chIP-chip arrays tailored to particular genomic regions. NimbleGen is also pushing the idea of array re-use, and currently guarantees three-time use of its chIP-chip products. "Our goal is to have a platform that is ten times re-usable in the next cycle," says Nuwaysir.

Agilent has also thrown its hat into this particular ring and launched its own chIP-chip platform last month for analysing activity at regulatory genomic regions. In the future it expects to offer a variety of custom and catalogue ChIP-chip products, as the company acquired Computational Biology of Cambridge, Massachusetts, in January. This firm was founded by Richard Young of the Whitehead Institute at the Massachusetts Institute of Technology in Boston and has key intellectual property in this area. In a recent publication in *Cell*, Young's group used a chIP-chip approach to produce high-resolution genome-wide maps of acetylation and methylation in yeast.

Splice-variant analysis

Microarray analysis may have produced a mountain of gene-expression data but the contribution of alternative splicing has largely been overlooked until now. The market for

former senior executives from the diagnostics industry. "We want to capture revenue not just from the chip but from the test itself," says Michael Tognotti, vice-president of sales and marketing at CombiMatrix.

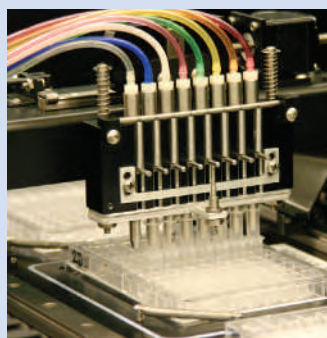
"We didn't feel that high-density arrays are the way to go for clinical utility," says Graham Lidgard, senior vice-president of research and development at Nanogen of San Diego, California. The company is targeting the clinical and diagnostic market with its NanoChip electronic microarray system. "If you want to do one or two targets, then reverse transcription-PCR is really the way to go, but if you want to look at four or five single-nucleotide polymorphisms to 100 targets across a number of patients, our technology is probably best suited to that application," Lidgard says.

The company's NanoChip 400 platform for multiplexed mutation

detection, out later this year, features increased automation and will offer a 400-site cartridge for increased throughput; the original version has 100 electrodes, or test sites.

An equity investment in Jurilab of Kuopio, Finland, announced in July, gives Nanogen certain rights to develop diagnostic products based on genes and gene markers discovered by Jurilab. The initial focus will be on cardiovascular and metabolic diseases, including adult-onset diabetes.

BioArray Solutions of Warren, New Jersey, has developed a novel twist on traditional bead-based systems where beads displaying proteins or DNA capture probes are in suspension and assays are performed 'in-tube'. BioArray's BeadChip format provides a bead array in the form of a single pre-assembled layer of colour-coded beads in random configuration on a small silicon chip. The company



Microfluidics: MetriGenix's TipChip platform for RNA and DNA.

is initially developing BeadChip assays to screen carriers for CF and Tay-Sachs disease, as well as for transfusion and transplantation medicine.

"It's a very diverse set of applications that we've already developed and deliver on a complete platform," says Michael Seul, president of BioArray Solutions. "As a small company, we've chosen to focus on certain

areas where we can have a large impact because they have been largely unexplored."

MetriGenix of Toronto, Canada, is developing genomic biomarker panels and genomics instrumentation for the diagnostic market. In the company's Flow-thru chip technology, molecular interactions occur within 3D microchannels rather than on a planar surface.

The company's TipChip technology is provided as part of the MGX8 automated system, which can process eight 1 × 1-cm silicon TipChips in parallel with up to 1,600 biomolecular targets or 'spots' per chip. The system is suitable for RNA, DNA or protein targets.

Sample preparation is still a time-consuming process in all platforms. Further automation on the front end would improve robustness and reduce costs.

METRIGENIX

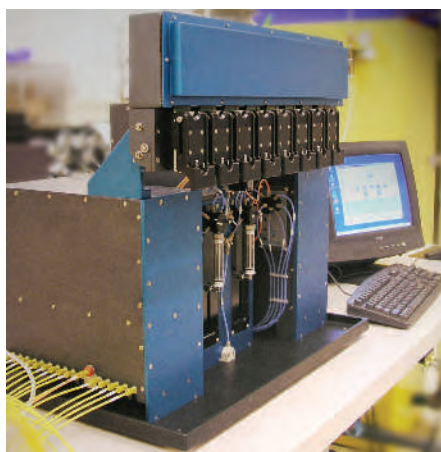
D.G.

splice arrays is now heating up, and companies such as ExonHit of Paris, France, and Jivan Biologics of Berkeley, California, are focused entirely in this area.

Most human genes are thought to undergo alternative splicing, the process by which individual genes can produce multiple messenger RNA and protein products. These so-called splice isoforms can have different, and even opposing, functions in the cell. With the recent commercial availability of splice-variant arrays, traditional gene-by-gene studies of alternative splicing are now taking a back seat to genome-wide methods. Further study in this area should provide a better — and more complete — understanding of the functional relevance of splice variants and of disease mechanisms.

ExonHit offers its SpliceArrays on a service basis, but from this September they can also be purchased directly from Agilent. Validation of the ExonHit approach, which involves the use of both exon- and junction-derived probes to follow every splice event associated with a given gene, was featured in a recent publication in *Nucleic Acids Research*.

The company offers both custom and catalogue SpliceArrays focused on specific gene families that are of particular therapeutic interest — cytokine and apoptosis pathways, nuclear receptors and co-regulators, G-protein-coupled receptors (GPCRs) and ion channels. “We are also in the process of expanding our line into mouse,” says Laurent Bracco, ExonHit’s vice-president of research and technology development, with products



Desktop DNA synthesis from CombiMatrix.

expected around the turn of the year.

Jivan has been working in this area for the past four years and offers its suite of TransExpress splice-variant microarrays, which are manufactured by Agilent. The company’s products include a genome-wide splice-variant array and a suite of arrays focused on gene families such as cytochrome P450, GPCRs, ion channels, kinases, phosphatases, phosphodiesterases, proteases and proteinases. Unlike ExonHit, Jivan’s splice arrays do not include exon probes. “Exon probes are neither necessary nor sufficient to measure RNA splicing and so we do not include them in our TransExpress product line. Using junction probes, however, all splicing events can be detected,” says Jonathan Bingham, Jivan’s chief technol-

ogy officer. The company also makes arrays to user specifications, and now offers a splice-variant microarray service through MOgene of St Louis, Missouri. “The next product will be an oncology array consisting of about 2,000 genes implicated in cancer along with their splice variants,” says Bingham.

At the high-density end of the market, earlier this month Affymetrix rolled out its exon-only array with a million exons on a single chip. This Human Exon 1.0 ST (sense target) array offers whole-genome, exon-level expression profiling on a single array and includes all annotated exons, as well as computationally predicted and empirically identified exon content. As such, the company says, researchers will be able to study known splicing events, as well as discover novel splice variants. GeneChip mouse and rat exon arrays are planned for later this year.

Microarrays are now a widely used technology for studying gene expression and regulation on a global scale and at high-throughput. Other applications for this technology include single-nucleotide polymorphism discovery and validation (see *Nature*, **422**, 917–922; 2003) and comparative genome sequencing. The use of microarrays is also moving downstream as the first microarray-based diagnostic tests begin to trickle into the marketplace (see ‘Microarrays move downstream’, page 1196). The future indeed looks bright. Has it really been only ten years? ■

Diane Gershon is assistant editor, *Nature Medicine*, Technical Reports

ON THE HARDWARE FRONT

Researchers interested in printing their own microarrays in-house now have an array of options — from contact (pin or split pin) and non-contact printing methods (bubblejet or piezoelectric inkjet) to *in situ* array synthesis.

Bio-Rad Laboratories of Hercules, California, launched the BioOdyssey Calligrapher miniarrayer last month, which prints oligos, proteins or cell lysates onto slides, membranes or 96-well plates. This bench-top instrument has an eight-pin print head and prints up to 16 slides at a time. Optional extras include humidity and temperature controls. Telechem International of Sunnyvale, California, has also added humidity controls to its NanoPrint microarrayer.

A plate-arraying option is now available on the BioRobotics MicroGrid ii and GeneMachines

OmniGrid Accent platforms from Genomics Solutions of Holliston, Massachusetts. This enables DNA and protein microarrays of up to 1,000 features per well to be printed in 96- and 384-well microplates, providing high-throughput, multiplexing capabilities for diagnostic, ELISA and protein-protein interaction applications.

Non-contact printers may have got a bad rap in the early days but Arrayjet of Mayfield, UK, hopes to win over researchers with its new Aj120 piezoelectric inkjet printer, which it launched in July. “Contact printing can affect the quality of the data you get because it introduces surface-based artefacts into the microarray,” says Duncan Hall, sales and marketing director for Arrayjet. Pin printers can also be very susceptible to atmospheric conditions, he says, particularly



Arrayjet makes non-contact printing high-throughput.

when hygroscopic printing buffers are used. “We carry our sample inside the print head so there isn’t any evaporation from the print head.”

The Aj120 includes a microplate stacker and lid lifter, which enables walk-away printing of microarrays from up to 48 (96- or 384-well) microtitre plates. With

the standard 12-sample connector block, 100 slides (samples and replicates side-by-side) can be printed in 90 seconds, says Hall. Additional replicates can be printed on the fly in no extra time. The Aj120 prints DNA, proteins and intact cells with a 100- μ m spot diameter. PerkinElmer Life and Analytical Sciences of Boston, Massachusetts, also offers contact and piezoelectric non-contact systems.

For those with deeper pockets, CombiMatrix sells a CustomArray desktop DNA synthesizer for the *in situ* synthesis of oligonucleotides on microarrays with up to 12,000 features. Synthesis occurs on a blank CombiMatrix semiconductor chip using standard phosphoramidite chemistry methods.

D.G.

COMPANY	PRODUCTS/ACTIVITY	LOCATION	URL	
Microarrays, microarray equipment and software				
Affymetrix	GeneChip arrays and reagents, instrument systems and software	Santa Clara, California	www.affymetrix.com	●
Agilent	Instrumentation for the life sciences, microarray equipment, custom and catalogue DNA microarrays, microarray kits	Palo Alto, California	www.agilent.com	
Alpha Innotech	AlphaScan microarray scanner,	San Leandro, California	www.alphainnotech.com	
Alpha Technology	Customized low- and medium-density DNA microarrays	Hürth, Germany	www.alpha-tec.net	
Applied Precision	arrayWoRx biochip and microplate readers	Issaquah, Washington	www.appliedprecision.com	
ArrayJet	AJ100 microarray inkjet spotter	Edinburgh, UK	www.arrayjet.co.uk	
BD Biosciences	Atlas arrays, labelling kits, AtlasImage 2.01/AtlasNavigator 1.0 software	Franklin Lakes, New Jersey	www.bd.com	
BioAutomation	MerMade DNA/RNA synthesizer, MagnaSpotter microarrayer	Plano, Texas	www.bioautomation.com	
BioAware	BioArchitech for microarray design	Hannut, Belgium	www.bio-aware.com	
BioCat	Arrays, SNP genotyping kits and services	Heidelberg, Germany	www.biocat.de	
Biocomputing Platforms	Data-management systems for genotyping and phenotype data	Espoo, Finland	www.biocomputing.fi	
BioDiscovery	CloneTracker, ImaGene 4.2, GeneSight 3.0, GeneDirector for microarray research and analysis	El Segundo, California	www.biodiscovery.com	
Bioforce Labs	NanoArrayer for protein and DNA microarrays	Ames, Iowa	www.bioforcelab.com	
BioMicro	MAUI microarray mixer chambers and hybridization chambers	Salt Lake City, Utah	www.biomicro.com	
Bio-Rad	Products, instruments and software for life-sciences research, BioOdyssey Calligrapher microarrayer	Hercules, California	www.bio-rad.com	
Clondia	DNA array technology (ArrayTube), PARTISAN microarray LIMS	Jena, Germany	www.clondia.com	
CombiMatrix	Customized DNA microarrays, RNAi	Mukilteo, Washington	www.combimatrix.com	
DNA Star	Desktop DNA and protein-sequence analysis software, ArrayStar for microarray analysis	Madison, Wisconsin	www.dnastar.com	●
Eppendorf	DualChip microarrays for human and mouse, reagents and consumables	Hamburg, Germany	www.eppendorf.com	
Eurogentec	Custom DNA microarrays	Seraing, Belgium	www.eurogentec.com	
Exiqon	Microarray slides and chips, LNA miRNA probes	Vedbaek, Denmark	www.exiqon.com	●
Fujifilm	Imaging systems for protein electrophoresis and array analysis	Stamford, Connecticut	www.fujimed.com	
GE Healthcare	Instruments, equipment and products for genomics, proteomics and cellular assays, MegaBACE capillary-based, high-throughput automated sequencing and genotyping systems, CodeLink platform for microarrays	Little Chalfont, UK	www.amersham.com	
Genetix	Picking, arraying, gridding, replicating, rearraying and microarraying robots for genomics and proteomics	New Milton, UK	www.genetix.co.uk	
Genomic Solutions	GeneTAC biochip system/GenoMAP pre-printed arrays, Cartesian Technologies products, BioRobotics products, GeneMachines products	Ann Arbor, Michigan	www.genomicsolutions.com	
GenoSensor	GenoExplorer range of microarrays and services	Tempe, Arizona	www.genosensorcorp.com	
Improved Outcomes Software	GeneLinker software for gene-expression analysis and proteomics	Inverary, Ontario	www.improvedoutcomes.com	
Invitrogen	Kits and reagents for cloning, genomics, proteomics, molecular and cell biology, microarrays and protein arrays	Carlsbad, California	www.invitrogen.com	
Illumina	BeadArray system for large-scale genotyping, custom oligonucleotide synthesis	San Diego, California	www.illumina.com	
Kreatech Biotechnology	Universal labelling system (ULS) for nucleic acids and proteins	Amsterdam, The Netherlands	www.kreatech.com	
LC Sciences	Oligonucleotides, DNA and peptide microarrays, miRNA detection services	Houston, Texas	www.lcsclences.com	
Luminex Corporation	xMAP platform for microbead-based multiplex assays	Austin, Texas	www.luminexcorp.com	
Memorec Stoffel	PIQOR microarrays, microarray services	Cologne, Germany	www.memorec.com	
MiraBio	Bead-based array systems and software	Alameda, California	www.miraibio.com	
Molecular Devices	GenePix scanners for microarray analysis	Sunnyvale, California	www.moleculardevices.com	
Nanogen	NanoChip automated workstation for genomics and proteomics analysis, electronic microarray chips, NanoChipR400 for diagnostic testing and diagnostic development	San Diego, California	www.nanogen.com	
NimbleGen Systems	Custom and catalogue microarrays, array CGH, ChIP-chip analysis, comparative genome sequencing, expression analysis	Madison, Wisconsin	www.nimblegen.com	
Panomics	DNA arrays for transcription regulation and cell signalling	Redwood City, California	www.panomics.com	
PerkinElmer Life Sciences	Micromax high-throughput human DNA microarrays, HydroGel BioChip for DNA and protein arrays, ProScan microarray analysis software	Boston, Massachusetts	las.perkinelmer.com	
Phalanx	Whole human genome DNA microarray	Hsinchu, Taiwan	www.phalanx.com.tw	
Premier Biosoft	Desktop bioinformatics packages for sequence analysis, Array Designer for microarray oligonucleotides	Palo Alto, California	www.premierbiosoft.com	●
Radius BioSciences	3XVP arrayer and other automated microarray equipment	Medfield, Massachusetts	users.rcn.com/radius.ma.ultranet	
Rosetta Biosoftware	Rosetta Resolver gene-expression data-analysis system	Seattle, Washington	www.rosettatabio.com	
Rosetta Genomics	miRNA microarray development	Rehovot, Israel	www.rosettagenomics.com	
Scanalytics	MicroArray Suite software for gene-expression analysis	Rockville, Maryland	www.scanalytics.com	
Scandinavian Micro Biodevices	Microfluidics device manufacturer	Lyngby, Denmark	www.smb.dk	
Schott Nexterion	Coated slides and reagents for spotting microarrays	Jena, Germany	www.schott.com	
SciGene	Little Dipper microarray processing system, microarray equipment	Sunnyvale, California	www.scigene.com	
Strand Genomics	Oligonucleotide design and microarray-analysis software	Bangalore, India	mail.strandgenomics.com	
Stratagene	Tools and reagents for molecular biology, genomics and proteomics, products for microarray construction and analysis	La Jolla, California	www.stratagene.com	●
SuperArray Bioscience Corporation	GEArray systems for pathway-specific gene-expression profiling	Frederick, Maryland	www.superarray.com	
TeleChem International	ArrayIt brand products for microarray analysis	Sunnyvale, California	www.arrayit.com	
Vysis	Microarrays, microarray readers and reagents	Downer's Grove, Illinois	www.vysis.com	
Whatman/Schleicher & Schuell	FASTSlides, FastQuant, ArrayVision FAST software for protein microarrays	Brentford, UK	www.whatman.com	
Zeptosens	SensiChip DNA microarray systems	Witterswil, Switzerland	www.zeptosens.com	

COMPANY	PRODUCTS/ACTIVITY	LOCATION	URL
In vitro diagnostics and genetic testing			
Agendia	Gene-expression analysis-based diagnostics for oncology	Amsterdam, The Netherlands	www.agendia.com
Asper Biotech	Genetic testing	Tartu, Estonia	www.asperbio.com
BioArray Solutions	Genetic typing and <i>in vitro</i> diagnostics	Warren, New Jersey	www.bioarrays.com
bioMérieux	<i>In vitro</i> diagnostics for medical and industrial applications	Marcy l'Etoile, France	www.biomerieux.com
Lambda	Biochips for genotyping and SNP detection	Freistadt, Austria	www.lambda.at
Metrigenix	Flow-Thru-Chip technology for genomic testing	Gaithersburg, Maryland	www.metrigenix.com
Roche Diagnostics	<i>In vitro</i> diagnostics	Indianapolis, Indiana	www.roche-diagnostics.us
Tm Bioscience	Genetic testing, Tag-It mutation detection kits, Tm100 Universal Array	Toronto, Canada	www.tmbioscience.com
VBC-GENOMICS Research	Microarray-based diagnostics	Vienna, Austria	www.vbc-genomics.com
Veridex	<i>In vitro</i> diagnostic oncology products	Warren, New Jersey	www.veridex.com
Services			
Enzo Life Sciences	Labelling and detection reagents for molecular biology research, including microarrays	Farmingdale, New York	www.enzolifesciences.com
Eppendorf Array Technology	Custom DNA microarrays	Namur, Belgium	www.eppendorf.com/eat/en/index.php
ExonHit	Drug discovery based on analysis of alternative splicing	Paris, France	www.exonhit.com
Genaco	Proprietary platform mirMASA for miRNA expression profiling using Luminex xMAP technology	Huntsville, Alabama	gene.genaco.com
Genisphere	3DNA probes for microarrays, microarray services, RNA amplification.	Hatfield, Pennsylvania	www.genisphere.com
GenScript	Kits and services for cell and molecular biology	Piscataway, New Jersey	www.genscript.com
Jivan Technologies	Gene Family Splice Arrays for analysis of alternative splicing	Berkeley, California	www.jivanbio.com
MOGene	DNA microarray services	St Louis, Missouri	www.mogene.com
Operon	Oligonucleotides, Oligo MicroArray Database	Huntsville, Alabama	www.operon.com
Orchid Cellmark	Array-based SNP genotyping and DNA profiling	Princeton, New Jersey	www.orchid.com
Oxford Gene Technology	Microarray services, custom Chip-chip arrays	Oxford, UK	www.ogt.co.uk
Pamgene International	5D-Pulse platform for DNA, protein and peptide microarrays	's-Hertogenbosch, The Netherlands	www.pamgene.com
Paradigm Array Labs	Microarray services	Research Triangle Park, North Carolina	www.icoria.com
Perlegen Sciences	Microarray-based SNP genotyping services	Mountain View, California	www.perlegen.com
General			
Agencourt Bioscience Corporation	SPRI (solid-phase reversible immobilization) products and instruments for DNA purification, plasmid purification kits, clones, genomics services	Beverly, Massachusetts	www.agencourt.com
Ambion	Silencer range of kits, vectors, siRNAs and reagents for RNAi, kits and products for RNA synthesis, isolation, measurement and analysis	Austin, Texas	www.ambion.com
Applied Biosystems	DNA sequencers, workstations and automated instruments for genomics and proteomics	Foster City, California	home.appliedbiosystems.com
Beckman Coulter	Automated tools for molecular biology, biochemistry, genomics and proteomics	Fullerton, California	www.beckmancoulter.com
BIOGENES	Customized monoclonal and polyclonal antibodies, immunoassays, peptides	Berlin, Germany	www.biogenes.de
Bioslide	Laboratory glassware, coated slides for microarrays, microscope slides	Walnut, California	www.bioslide.com
BMG LABTECH	Microplate and array readers and handling systems	Offenburg, Germany	www.bmg-labtechnologies.com
Cambrex	Gels and general products for molecular and cell biology research	Walkersville, Maryland	www.cambrex.com
Corning	Coated slides for microarrays, hybridization chambers	Acton, Massachusetts	www.corning.com
Epicentre	TargetAmp RNA amplification kits, ArrayPure RNA purification kits	Madison, Wisconsin	www.epibio.com
Hamilton Company	Microlab automated liquid-handling workstations	Reno, Nevada	hamiltoncomp.com
Integrated DNA Technologies	Custom oligonucleotides including locked nucleic acids phosphorothioate antisense oligonucleotides	Coralville, Iowa	www.idtdna.com
Leica Microsystems	DN digital microscope range, microscopes and microAdvertiserscope systems	Wetzlar, Germany	www.leica-microsystems.com
NuGEN Technologies	Ovation RNA amplification process	San Carlos, California	www.nugenttechnologies.com
Pepscan Systems	Custom and catalogue PepChip kinase substrate peptide arrays for kinase profiling, services	Lelystad, The Netherlands	www.pepscan.com
Perseus Proteomics	Antibodies	Tokyo, Japan	www.ppmx.com
Pierce Biotechnology	Components and consumables for molecular biology	Rockford, Illinois	www.piercenet.com
Promega	Vectors, reagents and kits for genomics, proteomics and cellular analysis, READ-IT system for SNP genotyping and mutation detection	Madison, Wisconsin	www.promega.com
Qiagen	Laboratory workstations and equipment for nucleic acid, protein and cell purification, reagents, kits and instruments for genomics, proteomics, molecular and cell biology	Valencia, California	www.qiagen.com
R&D Systems	Fluorokine MAP, for multiplexed cytokine and MMP assays	Minneapolis, Minnesota	www.RnDSystems.com
Roche Applied Science	Reagents and kits for molecular biology, functional genomics and proteomics research	Indianapolis, Indiana	www.roche-applied-science.com
Sigma-Aldrich	Reagents and biochemicals for life-sciences research, including slides and reagents for DNA microarrays	St Louis, Missouri	sigma-aldrich.com
Tecan	Genesis and Freedom EVO liquid-handling workstations and automated solutions for genomics	Männedorf, Switzerland	www.tecan.com
	LS series laser scanner		
SurModics	3D-Link activated slides for DNA microarrays	Eden Prairie, Minnesota	www.surmodics.com

- FOCUS
- SPOTLIGHT
- RECRUITMENT
- ANNOUNCEMENTS
- EVENTS

naturejobs

The path to success

Dario Alessi, winner of this year's EMBO Gold Medal, attributes his success to both his working environment and his non-traditional career choice. The medal, which is awarded by the European Molecular Biology Organization, recognizes outstanding research by young molecular biologists working in Europe.

Alessi works on cell signalling at the MRC Protein Phosphorylation Unit at the University of Dundee, Scotland. At the age of 37, he has already published more than 100 peer-reviewed papers — a fact that helped secure his award earlier this month. He says that his research unit in Dundee is an ideal place for young scientists to excel.

As he is funded by the UK Medical Research Council, Alessi has no teaching duties and writes very few grant applications. His unit hosts 80 research groups and 150 scientists in an 'open plan' research environment. "In all the labs, all the researchers interact with each other," Alessi says. "All the equipment is communal. If someone discovers something, it is transmitted very quickly throughout the building."

Alessi was attracted to Dundee after doing PhD work in London and Birmingham. And, contrary to conventional wisdom, he decided to stay on as a principal investigator after he completed his postdoc there. "Everyone told me not to do that, because everyone will associate you with your mentor," Alessi says. But staying on allowed him to maintain his momentum and he soon established his own reputation.

Alessi's success illustrates two scientific career points. First, working environment can be more important than university or laboratory pedigree — Alessi knows that Dundee isn't held in the same esteem as Cambridge or Harvard, but he doesn't much care. And sometimes you have to ignore professional dogma. Perhaps other young scientists should follow Alessi's trail — and maybe they too will achieve similar success before they turn 40.



Paul Smaglik, Naturejobs editor

CONTACTS

Publisher: Ben Crowe
Editor: Paul Smaglik
Marketing Manager: David Bowen

European Head Office, London
 The Macmillan Building, 4 Crinan Street
 London N1 9XW, UK
 Tel: +44 (0) 20 7843 4961
 Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com

Naturejobs Sales Director:
 Nevin Bayoumi (4978)
European Sales Manager:
 Andy Douglas (4975)

Natureevents: Sille Opstrup (4994)
UK/RoW/Ireland/Italy:
 Nils Moeller (4953)
 Irene Viglia-Atton (4944)
Scandinavia/Spain/Portugal:
 Evelina Rubio Håkansson (4973)
France/Switzerland/Belgium:
 Amelie Pequignot (4974)
Germany/Austria/The Netherlands:
 Reya Silao (4970)

Advertising Production Manager:
 Billie Franklin
 To send materials use London
 address above.
 Tel: +44 (0) 20 7843 4814

Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com
Naturejobs web development: Tom Hancock
Naturejobs online production: Niamh Shields

European Satellite Office
Germany/Austria/
The Netherlands:
 Patrick Phelan
 Tel: +49 89 54 90 57 11
 Fax: +49 89 54 90 57 20
 e-mail: p.phelan@nature.com

US Head Office, New York
 345 Park Avenue South,
 10th Floor, New York,

NY 10010-1707
 Tel: +1 800 989 7718
 Fax: +1 800 989 7103
 e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo
 Chiyoda Building,
 2-37 Ichigayatamachi,
 Shinjuku-ku,
 Tokyo 162-0843
 Tel: +81 3 3267 8751
 Fax: +81 3 3267 8746
Asia-Pacific Sales Director: Rinoko Asami
 e-mail: rasami@naturejpn.com

You say genomics, I say genetics...

Whatever title is afforded today's practitioners of molecular genetics, one thing is certain — success hinges on having a wide set of skills.

Ricki Lewis reports.

Training as a geneticist or genomicist today involves mastering a lot more than the mendelian ratios, central dogma and population equations of times past. Not only has the conceptual focus shifted from genes to genomes, but that change has added the new realm of bioinformatics. As a result, today's genetic scientists have skill sets that dwarf those of their mentors.

"Back in my day, the ability to manipulate a large amount of data was only for population geneticists. No longer. Molecular geneticists do this today too," says Joann Boughman, executive vice-president of the American Society of Human Genetics in Bethesda, Maryland.

Over the past few years, familiarity with, if not expertise in, bioinformatics has become not just desirable, but expected. "Classical and molecular genetics are still part of the training, but it is hard to believe that anyone with a PhD in genetics nowadays would not also have a background in bioinformatics," says Richard Gregory, director of research at Genzyme in Cambridge, Massachusetts.

After half a century of single-gene quests, the ability to sequence genomes has reframed research in the much broader context of gene interactions. But the distinction between geneticist and genomicist remains fuzzy and perhaps artificial, for the two titles represent a continuum that differs more in scope than skills. "I think of genetics as disease-gene hunts," says Claire Fraser, president and director of The Institute for Genomic Research (TIGR) in Rockville, Maryland. "That's never been what we do at TIGR. We look at whole genomes and do comparative analyses." Since TIGR's debut in 1992, researchers there have sequenced the genomes of more than 200 species.

Growth factors

Whether the job is to investigate the 1,200 variations on the cystic fibrosis gene theme, or the billions of bases of the wombat genome, the challenge when the sequencing and assembling ends is the same: how to extract the most useful information from the data. Colin Hill, chief executive of the biotech company Gene Network Sciences in Ithaca, New York, fittingly calls the new breed of scientist who can do this a "true hybrid".

For those who can both churn out data and analyse them in a way that solves problems, career opportunities are diverse and growing. They range from working with families who have a genetic disease, to developing strategies to control insect pests, to identifying disaster victims, to improving the pop in popcorn.

Whatever the new hybrid eventually comes to be called, the job market is expanding fast. For example, the Austrian Genome Research Programme is

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS



Colin Hill: the new breed of scientist is a true hybrid.

Out of the blue: jobs are being created at a startling rate.

beginning phase two of its nine-year effort to foster public-private partnerships, with a goal of launching 30 biotech and genomics companies. The programme anticipates that the number of jobs in genomics will quintuple in the next decade.

On a more global level, there is the special programme for research and training in tropical diseases, or TDR: a collaboration among UNICEF, the World Health Organization (WHO), the United Nations and the World Bank. TDR seeks expertise in parasite genomes and bioinformatics as part of its new 'portfolio of putative drug targets' programme, according to Solomon Nwaka, manager of genomic and discovery research at the WHO in Geneva. And at the International Potato Center in Lima, Peru, a new programme that embraces genomics will establish links between agriculture and health.

Many academic institutions and medical centres are spinning off programmes with a genomics/bioinformatics bent. Cornell University in Ithaca, New York, seeks cell biologists with a genomics approach for a new institute, and Albert Einstein College of Medicine in New York is recruiting faculty members for its new Integrated Genetics and Systems Biology Initiative. At the Beatson Institute for Cancer Research in Glasgow, Scotland, established research groups are growing and new ones forming in vast and expanding areas such as the control of gene expression, which uses microarray technology to track protein production in cells.

Undergraduate and graduate education recognize the new hybrid too. At Cook College in New Brunswick, New Jersey, the biotechnology programme integrates genetics with the fields it has spawned.

"Students take biochemistry, organic chemistry, genetics both classical and molecular, and also bioinformatics, biostatistics and bioethics," says Cook curriculum coordinator Barbara Zilinskas. "We also offer courses in regulatory matters, intellectual property and communications." Graduates have had remarkable success in the job market. "It is a good time for young people with a BS or PhD, who are smart and well-trained, to secure jobs in biotechnology, particularly related to human medicine," she adds. At the graduate level, the University of Pennsylvania offers a PhD in genomics and computational biology.

The picture isn't quite so rosy for agricultural biotechnology, perhaps because of the movement against genetically modified organisms, Zilinskas says. Still, Syngenta, one of the half-dozen large companies that dominate agricultural biotechnology, is hiring plant geneticists, according to Anne Burt, spokeswoman for the company, based in Basel, Switzerland. And the UK agricultural group Cargill of Cobham, Surrey, seeks scientists with PhDs in molecular genetics, genomics or related fields to identify valuable gene variants and introduce them into crop plants.

Eclectic opportunities

Non-academic jobs in genetics have traditionally been in the healthcare and agricultural sectors, but the evolution to genomics has expanded those areas and catalysed the formation of new ones. For example, a clinical geneticist requires an MD and residency, but PhD scientists will increasingly be tapped to educate the medical community, says Boughman. Family doctors, paediatricians and specialists in internal medicine need to understand genetic predispositions in a different way, and know more about genetics and calculating risk, she says.

Genetic-testing laboratories are experiencing a boom that reflects soaring public interest, thanks to awareness of genetic components of common disorders, a growing interest in family history, and widespread use of forensic tests in the aftermath of disasters. Myriad Genetics in Salt Lake City, Utah, for example, hires research associates with bachelor's or master's degrees to isolate, amplify and sequence DNA for genetic and forensic tests. Demand is so high that they run 'graveyard shifts' overnight for those who don't mind extracting DNA in the wee hours. For scientists with superb people skills, another position is women's health account executives. They educate physicians about the company's tests for inherited cancers. And Sorenson Genomics, also of Salt Lake City, has spun off business units to satisfy the public's thirst for information — GeneTree to track relationships and ethnicity, and Relative Genetics for genealogy. Both target the consumer, says Terry Carmichael, vice-president of marketing and sales. The company hires MS and PhD geneticists.

Genetics/genomics transcends humans, and that's where the need for the new hybrid is greatest — comparative genomics, the *raison d'être* at TIGR. Fraser has witnessed TIGR's workforce evolve over the past decade. "Early on we had a small cadre of people who built our bioinformatics capability, and they trained everyone. That's not the case anymore. People



Julian Parkhill: the most important aspect is biological interpretation.

come in with that level of expertise, because they've had to use it in their training," she says.

At the Pathogen Sequencing Unit within the Sanger Institute in Hinxton, Cambridge, UK, scientists at all levels are also highly trained. Technical staff with biology BScs sequence DNA and manipulate data, and PhD-level staff handle the more detailed molecular biology and troubleshooting, says senior investigator Julian Parkhill. Yet expertise in traditional biology is still highly valued — PhDs with postdoc experience in microbiology or parasitology tackle data analysis. "This is necessary because the most important aspect is the biological interpretation of the data in the context of the organism," Parkhill adds.

Beyond biology

A degree in a subject beginning with 'gen-' isn't required to mine genome data. Physical scientists and the computationally inclined are in demand too, taking things full circle to the core of physicists who founded molecular genetics half a century ago. Gene Network Sciences has teams of molecular biologists, mathematicians, theoretical physicists and software engineers. "Drug discovery and development require the ability to process data from the human genome project," Hill explains. "A company like us focuses on computation, but the 'gene' in our name means we are still doing biology."

Expansion beyond biology is also evident in government-sponsored research. The new Chemical Genomics Center of the US National Human Genome Research Institute, for example, seeks chemists to develop small-molecule probes to investigate physiological processes revealed in genome sequences.

The distinctions and limits of the geneticist/genomicist hybrid have not yet fully crystallized. "No two people with a PhD in genomics have the same background, and that is representative of the field," says Fraser. Instead of trying to master the list of '-atics' and '-omics', she concludes, "focus on some aspect of this large, comprehensive field".

Ricki Lewis is a freelance writer in Scotia, New York.

A TURF WAR IN GENETIC COUNSELLING?

Genetic counsellors help families navigate the maze of genetic testing options. The field began in the 1970s with a focus on rare, single-gene disorders. Since then, genetic counselling, like genetics itself, has diverged and diversified to include more common disorders. But territoriality has emerged, at least in the United States, where most of the 2,200 genetic counsellors have master's degrees in the field and are certified by the American Board of Genetic Counseling.

Because certification isn't required to practice, nurses, social workers and others do so, and some testing companies even replace flesh-and-blood counsellors with pamphlets, videos and websites. That worries Kelly Ormond, who is president of the National Society of Genetic Counselors. "For complicated genetic test interpretation it may be useful to consult a genetic counsellor," says Ormond.

Part of the problem is an inaccurate image of the genetic counsellor's role, persisting from the 1970s. "The genetic counsellor has a very narrow scope of practice, specific to single-gene disorders," says Kathleen Calzone, a senior nurse specialist at the National Cancer Institute who helped develop genetics/genomics competencies and curricula for nurses. She adds that nurses who specialize in genetics offer what genetic counsellors do, and more besides.

But Bonnie Liebers, a board-certified senior genetic counsellor at Northeast Health in Albany, New York, disagrees. Specially trained in cancer genetics, she cites testing for the breast-cancer gene *BRCA1* as an example. "A patient may get a gene test, but not a four-generation pedigree to determine which of the dozen or so genes that predispose to breast cancer she might have."

R.L.

MOVERS

David Bentley, chief scientist, Solexa, Chesterford, UK



1993–2005: Head of human genetics and principal investigator, Wellcome Trust Sanger Institute, Hinxton, UK

1991–93: Senior lecturer, United Medical and Dental Schools of Guy's and St Thomas's Hospitals, London

1988–91: Lecturer, United Medical and Dental Schools, London

When David Bentley was in his early teens, he was already nurturing his fascination with molecular biology. It was not yet being taught in schools, so he learned about life at the molecular level at home with his mother, a biology teacher who shared his interest. Together, they were often up late at night reading, among other things, the early monographs on the structure of DNA.

This childhood excitement — and later, encounters with pioneering geneticists such as John Sulston — were the starting points for a career spent at the forefront of major advances in human genetics. With a PhD in molecular biology, Bentley was a postdoctoral fellow at Guy's and St Thomas's Hospital in London, studying the mutations that cause genetic diseases such as haemophilia. Drawn into the small but growing community of geneticists, he began meeting Sulston, who became an important mentor. It was he who invited Bentley to be a founding member of the Sanger Centre (now the Wellcome Trust Sanger Institute) near Cambridge, UK, one of the world's top genome centres and a major player in the Human Genome Project.

While at the Sanger, Bentley met two lecturers in the chemistry department of the University of Cambridge as they were founding their company, Solexa. He was captivated by their vision for a new kind of DNA technology: anchoring millions of single DNA molecules on a chip for simple, fast, cheap DNA sequencing and analysis. "It swept away so many problems that we had to deal with every day in the lab," says Bentley. He joined the new company's advisory board.

About four years ago, Bentley was tempted to join the company full-time, but decided to stay at the Sanger because he was starting new genetics projects there. He continued to be an adviser to Solexa as it moved towards its goal of building an instrument that can sequence an entire human genome at a fraction of today's costs. Over the past year and a half, it became clear that the technology was beginning to work, says Bentley, and he wanted to devote more time to applying it to human genetics.

Bentley is now moving to Solexa as chief scientist, which in many ways is sending him back to his early days in molecular biology. "It's refreshing to return to my roots," says Bentley. He doesn't see this as abandoning academia in favour of industry. "I see a coming together of academia and industry, not a split," says Bentley. "It's possible to bring an academic ethos to a commercial setting and to bring a commercial ethos to an academic setting."

Corie Lok

SCIENTISTS & SOCIETIES

An academy of their own

No time in a scientific career feels lonelier than the limbo reserved for postdocs on the bumpy road to tenure. This is particularly true in Germany, with its hierarchy that segregates the tenured 'haves' from the 'have-nots'.

The Young Academy, founded in 2000, each year brings ten young scientists out of this isolation. A joint venture of the Berlin-Brandenburg Academy of Sciences and the Leopoldina Society, it is funded by the German education and research ministry. Membership is for five years and is open to any German academic who has received a PhD in the previous three to seven years; selection aims to be purely merit-based. Members receive some grant money but are not guaranteed positions. The goals are modest: interdisciplinary exchange, outreach to society, and giving young scientists a unified voice. Yet, after five years, the results have been unique.

At a time when top-down reforms in European higher education and research are politically driven, the Young Academy stands out as a fruitful grassroots approach to bringing scientists from disparate fields together. Our activities have included launching the first large-scale study of the effects of Germany's reformed tenure system on career development and interdisciplinary projects involving,

for example, neuroscientists, psychologists and philosophers.

Implementing these activities was often laborious, but for many members, this was the first chance to interact with scientists outside our own fields. For me as a physicist, exposure to the conceptual sophistication of the humanities was an education in precision beyond formulae. I revelled in learning first-hand about current malaria research and the workings of bee vision; as a young scientist, I felt my voice was heard.

The Young Academy is too new and reaches too far beyond disciplinary borders to help much with individual career advancement. But its real *raison d'être* ranges beyond job opportunities: to help young scientists develop a fuller view of the scientific universe; to give this view human meaning by associating it with faces and friends; to remind them that science is an endeavour of freely organized minds; and to urge them to take this spirit to the scientific community at large.

I recommend other countries to follow this example. Young academies may be a way to give the old idea of scientific academies a second lease of life.

Ulrich Schollwöck is a professor of theoretical physics at RWTH Aachen University. He is a founding member and former president of the Young Academy.

GRADUATE JOURNAL

Eye-opening meeting

Sometimes we do not realize how much we benefit from seemingly insignificant events. I recently attended a conference where I presented a poster. At first, I didn't think this conference would be anything special; just part of the ordinary life of every scientist. But going to it turned out to be a really good decision.

I met several interesting scientists who became interested in my project. I received a lot of useful feedback on this work, which was of great value given the significant number of specialists from the field there. I learned some new techniques. And I realized how misguided my doubts about coming had been.

Attending the conference was important for me, not only scientifically but also socially. I went with some colleagues from the institute where I work. I cannot imagine a better opportunity to get to know your co-workers as real people than at a conference. Suddenly you see certain features that you would not observe in any other situation: you can see who are the true professionals.

I also made friends with young scientists from the United States and from some other European countries. We discussed how we could work together and share information to improve our skills. The conference was indeed a useful experience for me, especially as I am new in the field and am just starting to build my career. I will attend this kind of event more often.

Karolina Tkaczuk is a graduate student at the Technical University of Lodz, Poland.

Calculation Quest

And the winner is...

Andrew Crumey

Professor Beatrice Ghosian, flanked by two male colleagues, was about to announce the winner of *Calculation Quest 2005*.

"Ladies and gentlemen," she began, "as you know, we invited the global public to participate in a unique computational experiment. Simply by watching our series on their Personal Entertainment Consoles, viewers have donated a portion of processor time to be used on a massive calculation chosen by our panel of experts.

"Initially, 50 calculations were proposed by our team of cultural commentators, style gurus and celebrity mathematicians. Over the weeks, that list was reduced through a rigorous process of peer review, group challenges and luck, until just three were left.

"Number one: calculation of the complete field equation of non-perturbative M-theory. Number two: to find a solution to world poverty and inequality. Number three: to create a perfect work of art. I now hand you over to Zane Budolsky, professor of comparative studies at Pennine University, Alston."

Budolsky, seated beside her, wore a thirties-style jumpsuit and had the air of an old-fashioned television philosopher. "Let me elucidate the finalists' charms and negativities," he said. "First, the equation of M-theory. This has been an outstanding problem of theoretical physics for more than half a century. And during all that time, physicists have assured us it will be solved next week. So let's wait until next week: they can do it for themselves.

"Number two. Solution to world poverty. We really liked this. But as my colleague Professor 'Rude Ron' Ronson has said: if we can't work out the answer in our own heads, maybe we don't deserve a solution found by other means. And that leaves the third option..."

He was cut off by a wave of Beatrice Ghosian's jewelled hand. "But the suspense isn't over. In fact, it's only beginning. The panel also had to take account of the people's choice, and I can announce that in this category, after a tough contest, the winner was...to work out where flies go to in winter."

Warm applause from the audience — then another flick of Beatrice Ghosian's long-nailed fingers, commanding silence. "But it was not to be. I now give you Starkling Bland, emeritus professor of colour at Ocean University 3, St Helena, who will explain why."



JACEY

Slumped in his chair, Bland seemed soundly asleep, possibly working off the effects of a long night of narcotaine smoothies. Given a prod, he roused himself and cast his bleary eyes across the multitude.

"The calculation, Stark," Professor Ghosian prompted. "Perfect work of art."

"Total," he said, suddenly igniting. "What an exceptional field it's been in this year's *Calculation Quest*. Particularly as this is the only year the show has run. And I hear they've axed the next series. So this is it, folks: the one and only. And that's why we really craved the work-of-art thing, number three. Loved it. Great package. The bidding team was so good. That housewife from Frankfurt. And the little guy with the thing in his ear. Personalities. That's what it's all about. You want a really great calculation — it's all about selling yourself.

"But which work of art? A symphony for wind chimes? An animated movie? A make-over? Art's a big sausage. Here's what we did: we tossed a coin. And we decided on literature."

From the audience, a low but distinctly perceptible moan. Professor Ghosian's features registered the shift: *Calculation Quest* was in sudden danger of going all wrong. With the second series cancelled, Ghosian's academic career was already on the rocks. Now it was her pension that was at stake.

"Yes!" she cried with forced animation, silencing Starkling Bland at a stroke. "Literature. You may call it outmoded, but that's not how we see things at *Calculation Quest*. Throughout the past seven days, the Personal Entertainment Consoles of 3.1 billion

viewers have been at work. We fed in every literary masterpiece in history; every critical analysis and commentary; every review. We left processing power to do what no human has ever achieved, racing through a hundred lifetimes of development and experience, gleaned through the literature of ages. And now we have it — yes, here in this room. The computation ended today at 5:32 universal time. In front of me I have the perfect work of literature."

It was inside a box whose lid Professor Ghosian now opened. What came out was not a book, but an envelope.

"Micro!" laughed Starkling Bland. "We got ourselves a haiku!"

Beatrice Ghosian was unruffled. "It has long been known that poetry is the highest form of literature. And I don't suppose the field equation of M-theory would have filled a novel either." She tore open the envelope, pulled out a white card and stared at it in silence.

"Go on, then," said Zane Budolsky. "Let's hear it!"

All the literature in history had gone into this. And that, Beatrice Ghosian realized, was the problem. Somehow, all those computers had failed to add up, but instead had only averaged — like the show she'd devised and now saw sinking along with her hopes. There were two short words printed on the card. They summed up her feelings exactly.

Andrew Crumey did a PhD in theoretical physics at Imperial College London, and is currently literary editor of *Scotland on Sunday*. His latest novel is *Mobius Dick* (Picador).